Methuen's Monographs on Physical Subjects
General Editor: B. L. Worsnop, B.Sc., Ph.D.

COSMOLOGICAL THEORY

METHUEN'S MONOGRAPHS ON PHYSICAL SUBJECTS

General Editor: B. L. Worsnop, B.Sc., Ph.D.

APPLETON, SIR EDWARD	THERMIONIC VACUUM TUBES
ARNOT, F. L.	COLLISION PROCESSES IN GASES
BEECHING, R.	ELECTRON DIFFRACTION
BRAILSFORD, F.	MAGNETIC MATERIALS
CHAPMAN, S.	THE EARTH'S MAGNETISM
CHILDS, W. H. J.	PHYSICAL CONSTANTS
DINGLE, H.	THE SPECIAL THEORY OF RELATIVITY
EMELÉUS, K. G.	THE CONDUCTION OF ELECTRICITY THROUGH GASES
FÈVRE, R. J. W. LE	DIPOLE MOMENTS
FLINT, H. T.	WAVE MECHANICS
FRANCIS, V. J.	FUNDAMENTALS OF DISCHARGE TUBE CIRCUITS
FRASER, R. G. J.	MOLECULAR BEAMS
HAGUE, B.	AN INTRODUCTION TO VECTOR ANALYSIS FOR PHYSICISTS AND ENGINEERS
HIRSCHLAFF, E.	FLUORESCENCE AND PHOSPHORESCENCE
JACKSON, L. C.	LOW TEMPERATURE PHYSICS
JACKSON, L. C.	WAVE FILTERS
JACKSON, WILLIS	HIGH FREQUENCY TRANSMISSION LINES
JACOB, L.	HIGH VOLTAGE PHYSICS
JAMES, R. W.	X-RAY CRYSTALLOGRAPHY
JOHNSON, R. C.	ATOMIC SPECTRA
JONES, T. J.	THERMIONIC EMISSION
JOSEPHS, H. J.	HEAVISIDE'S ELECTRIC CIRCUIT THEORY
KNUDSEN, MARTIN	THE KINETIC THEORY OF GASES
LAMONT, H. R. L.	WAVE GUIDES
McCREA, W. H.	RELATIVITY PHYSICS
McVITTIE, G. C.	COSMOLOGICAL THEORY
MANN, W. B.	THE CYCLOTRON
OATLEY, C. W.	WIRELESS RECEIVERS
OWEN, D.	ALTERNATING CURRENT MEASUREMENTS
PORTER, A. W.	THE METHOD OF DIMENSIONS
PORTER, A. W.	THERMODYNAMICS
RATCLIFFE, J. A.	THE PHYSICAL PRINCIPLES OF WIRELESS
SAS, R. K., and PIDDUCK, F. B.	THE M.K.S. SYSTEM OF ELECTRICAL UNITS
SCHONLAND, B. F. J.	ATMOSPHERIC ELECTRICITY
SOMMER, A.	PHOTOELECTRIC CELLS
STONER, E. C.	MAGNETISM
STYLE, D. W. G.	PHOTOCHEMISTRY
SUTHERLAND, G. B. B. M.	INFRA-RED AND RAMAN SPECTRA
SUTTON, O. G.	ATMOSPHERIC TURBULENCE
TEAGO, F. J., and GILL, J. F.	MERCURY ARCS
TEAGO, F. J.	THE COMMUTATOR MOTOR
TEMPLE, G.	THE GENERAL PRINCIPLES OF QUANTUM THEORY
TOLANSKY, S.	HYPERFINE STRUCTURE IN LINE SPECTRA AND NUCLEAR SPIN
WHITE, F. W. G.	ELECTROMAGNETIC WAVES
WILLIAMS, W. E.	APPLICATIONS OF INTERFEROMETRY
WILSON, A. J. C.	X-RAY OPTICS
WORSNOP, B. L., and CHALKLIN, F. C.	X-RAYS

COSMOLOGICAL THEORY

by

G. C. McVITTIE
M.A., Ph.D., F.R.A.S.
Professor of Mathematics in the University
of London, Queen Mary College

METHUEN & CO. LTD.
36 ESSEX STREET W.C.
London

First published . . . . October 7th 1937
Second edition 1949

CATALOGUE NO. 4005/U

PRINTED IN GREAT BRITAIN

PREFACE

COSMOLOGICAL theory is that branch of physics which deals with the structure of matter in its most bulky and massive state, the whole physical universe being regarded as a single system whose broad features are to be investigated. The subject is necessarily highly mathematical, but, in this introductory account, attention has been concentrated on those developments most easily comparable with observation to the exclusion of others of a purely mathematical interest. The first four chapters are concerned with the observational material, with the tensor calculus and with the expanding universes of general relativity. In the last chapter, Professor E. A. Milne's kinematical theory of the universe is developed in a way different from that used by its distinguished author. The theory is thereby generalized to any Riemannian space and, at the same time, the points in which it resembles, as well as those in which it differs from, general relativity are perhaps more clearly brought out.

It is hardly possible to give detailed references to the three standard works on this subject to which every writer must of necessity be heavily indebted. Sir A. S. Eddington's *Mathematical Theory of Relativity* (2nd ed., 1924), Professor R. C. Tolman's *Relativity, Thermodynamics and Cosmogony* (1934) and Professor E. A. Milne's *Relativity, Gravitation and*

World-Structure (1935), do not therefore appear in the short bibliographies given at the ends of the chapters. These contain references to sources of additional information and to detailed investigations on special points. No attempt at compiling a comprehensive bibliography has, however, been attempted.

It is a pleasure to thank Professor G. Temple who kindly read the manuscripts and made many valuable suggestions and criticisms. To Sir A. S. Eddington and to Professor E. T. Whittaker I owe a no less heavy, if more general, debt. It is to their teaching that my acquaintance with cosmological theory is due and it is their ideas which have most deeply coloured my own outlook.

G. C. McV.

March 1937

PREFACE TO SECOND EDITION

THE opportunity has been taken, in reprinting the 1937 edition, of correcting a number of minor errors and misprints. Imperfect though the book may be, I have not attempted a major revision which could not profitably be made without observational data from the 200-inch telescope, whose construction has been so long delayed by the war.

G. C. McV.

December 1948

CONTENTS

CHAPTER I

THE EXTRA-GALACTIC NEBULAE

DURING the last thirty years two distinct generalizations of classical mechanics have been developed. On the one hand there is the quantum theory, concerned with atomic phenomena ; on the other hand, general relativity, describing the phenomena associated with bodies of large mass. Modern cosmological theory deals with the system of the extremely massive extra-galactic nebulae and its subject-matter is the structure of the universe as a whole. Having arisen as a branch of general relativity, cosmology has since reacted on mechanics and has suggested new generalizations of classical Newtonian theory differing from both general relativity and quantum theory.

THE EXTRA-GALACTIC NEBULAE

The name extra-galactic nebulae is given to certain faint telescopic objects some of which can be resolved into separately visible stars surrounding what appears to be a mass of glowing gas. These systems are usually of ' spiral ' form, consisting of a disk-like distribution of stars arranged in trailers or winding arms around the central gaseous core, the whole having the appearance of a rotating body. Amongst the nebulae in general, the spirals preponderate in the ratio of three to one, the rest being of spherical, elliptical, or irregular shape. It is believed that the latter types also consist of stars although no spherical or elliptical nebula has actually been resolved. Sometimes it happens that a spiral nebula is presented to us ' edge on '. It is then frequently observed that a thin band of obscur-

ing material occurs in the equatorial plane of the nebula, cutting off all the light from those parts of the nebula which lie behind it.

In spite of the appearance of rapid rotation presented by spiral nebulae, the angular velocities [1] of only three of them have proved measurable spectroscopically with any certainty.

STELLAR MAGNITUDES AND DISTANCES

The first problem connected with the extra-galactic nebulae is the determination of their distances. Fundamentally there is only one method available, viz. the identification in the nebula of stars of known brightness, or 'luminosity', and a comparison of the apparent with the true luminosities of such stars. As a preliminary it is therefore necessary to explain how the luminosity of a star is measured and how a knowledge of the luminosity leads to an evaluation of its distance.

The apparent brightness of a star may be measured by comparing it visually with an artificial star or with a set of fundamental stars. It is measured by a number called the apparent visual magnitude of the star and the scale of magnitudes is arranged so that a first-magnitude star is $\sqrt[5]{100}$ or 2·512 times as bright as a second-magnitude star and so on, the ratio between each successive magnitude being the same. The result of this system is that really bright objects have negative apparent magnitudes, the full moon being of apparent visual magnitude $-$ 12·5, Sirius $-$ 1·58, whilst the Pole Star is of apparent visual magnitude $+$ 2·0 approximately. The spiral nebulae with which we shall be concerned have apparent magnitudes lying between $+$ 4 and $+$ 21. Similarly a *photographic apparent magnitude* is defined by reference to the relative diameters, or the degrees of blackness, of the images of stars on the photographic plate when exposures of equal

so that stars of a certain spectral type have the same magnitude in each. Stars of other types then show systematic differences which need not concern us here. Throughout this book we shall understand all magnitudes to be photographic.

The apparent magnitude of a star must clearly depend on its distance as well as on its intrinsic luminosity. We therefore define an *absolute* magnitude as the apparent magnitude the star would have if it were placed at a distance of 10 parsecs from the Sun. This unit of distance is related to the more familiar light-year and to the ordinary units of distance by

$$\begin{aligned} 1 \text{ parsec} = 3{\cdot}258 \text{ light-years} &= 1{\cdot}92 \times 10^{13} \text{ miles} \\ &= 3{\cdot}08 \times 10^{18} \text{ cm.} \quad . \quad . \quad (1.1) \end{aligned}$$

A star distant one parsec would subtend an angle of one second of arc when viewed from the two extremities of a radius of the earth's orbit. We can relate the apparent magnitude, m, of the star with its absolute magnitude, M, and its distance in parsecs, D, as follows. Let E_m be the total quantity of energy received per second in a beam of light from a star and let E_0 be the corresponding quantity from a star of zero apparent magnitude. Let also E be the energy which would be received if the star had been at a distance of 10 parsecs. Then by the definition of apparent magnitude, we have

$$E_0/E_m = (2{\cdot}512)^m \quad . \quad . \quad . \quad . \quad (1.2)$$

and, by the inverse square law for the diminution of light intensity,

$$E_m/E = 10^2/D^2 \quad . \quad . \quad . \quad . \quad (1.3)$$

Hence from (1.2)

$$m = 2{\cdot}5 \log_{10} (E_0/E_m),$$

and

$$M = 2{\cdot}5 \log_{10} (E_0/E).$$

We therefore obtain

$$M - m = 2{\cdot}5 \log_{10} (E_m/E) = 5 \log_{10} (10/D)$$

and finally

$$\log_{10} D = 0{\cdot}2(m - M) + 1 \quad . \quad . \quad . \quad (1.4)$$

THE CEPHEID STARS

The formula (1.4), since it required a knowledge of the absolute magnitudes of the stars involved in the extra-galactic nebulae, was of little use for the determination of nebular distances until a remarkable discovery made in 1912 opened up the way to further advances. In the southern heavens there are two irregular nebulae, the Magellanic Clouds, which are easily resolved into stars by telescopes of moderate power. It was observed [2, 3] that they contained large numbers of a certain kind of variable star, called 'Cepheid' after the star δ Cephei of this character which belongs to the same nebula as does the Sun. The brightness of a star of this kind varies through a fixed range of magnitude with a fixed period and the rise and fall in brightness is of a peculiar and easily recognized kind. Observation revealed that the Cepheids in the Magellanic Clouds, whose periods ranged from fifteen hours to over one hundred days, exhibited a close correlation between the period of, and the average apparent magnitude during, the fluctuation of brightness. Since the two nebulae are very faint, it is clear that the stars they contain must all be at approximately the same very great distance from the Sun. Thus the differences of average apparent magnitude must be due to differences of intrinsic brightness. It follows that *the period of a Cepheid variable must be correlated with its average absolute magnitude.* Assuming that all Cepheids possess this property, the coefficients of the correlation can be fixed by calculating the distances of those Cepheids which are, astronomically speaking, near the Sun. It has been found possible to do this by methods independent of absolute-magnitude criteria. A 'period-luminosity' curve is then drawn in which absolute magnitude is plotted against period of light fluctuation. From this curve the absolute magnitude of any newly discovered Cepheid whose period has been determined can be read off. Its distance is then calculated by means of (1.4) from the observed apparent magnitude.

In addition to Cepheids, variable stars of other types have been found in the Magellanic Clouds. The absolute magnitudes of such stars are also roughly known and serve to check the distances found by the Cepheids. Some 200 stars of known absolute magnitude have thus been located in the Small Magellanic Cloud alone. Its distance from the Sun proves to be $2{\cdot}9 \times 10^4$ parsecs, whilst its companion, the Large Magellanic Cloud, is a little closer to us at $2{\cdot}62 \times 10^4$ parsecs.

Unfortunately, Cepheid stars have been found in only ten nebulae so far, including our two nearest neighbours, the Magellanic Clouds. The most distant nebula of this group, the nebula M 81, lies at $7{\cdot}3 \times 10^5$ parsecs. Two important deductions are made from these observations: (1) that the absolute magnitude of an average nebula is $-14{\cdot}2$; (2) that the absolute magnitude of the brightest stars involved in a nebula is always very nearly $-6{\cdot}12$.

NEBULAR DISTANCES [3, 10]

When individual types of stars can no longer be identified in the nebula but stars as such can still be seen, resort is made to the assumption that stars have a maximum luminosity. This is in accordance with present-day theories of stellar constitution and also with the empirical result found for the ten nearby nebulae. The distances of some 135 more nebulae have thus been calculated from the apparent magnitudes of their brightest stars, using (1.4) with $M = -6{\cdot}12$. From the distances of nebulae of this group and their apparent magnitudes it is again deduced that the average absolute magnitude of a nebula is $-14{\cdot}2$. The most distant nebulae to which this method has been applied are those belonging to a cluster in the constellation Virgo. They lie at $2{\cdot}34 \times 10^6$ parsecs, the error in this value being estimated at 15 per cent.

For still greater distances, when stars can no longer be seen separately in the nebula, the only available method is to assume that the average absolute magnitude of the nebula is $-14{\cdot}2$ and to deduce the distance directly from (1.4) using the apparent magnitude of the nebula itself.

This procedure is only applicable with any certainty to nebulae which, owing to their proximity and to their small spread in apparent magnitude, evidently form physically connected groups. The most frequent apparent magnitude in such a cluster of nebulae is then likely to correspond to absolute magnitude $-14{\cdot}2$. In twenty-five clusters of nebulae studied by Shapley the maximum distance was 33×10^6 parsecs.

The faintest nebulae visible with the 100-inch telescope at the Mount Wilson Observatory have apparent magnitudes of about $+21$. Their distances would thus be of the order of 110×10^6 parsecs.

THE GALAXY AND THE DIMENSIONS OF NEBULAE

The Cepheid period-luminosity curve also serves to indicate the dimensions of the Galaxy, the nebula to which the Sun himself belongs. This nebula contains all stars visible to the naked eye, the stars of the Milky Way, &c. The Galaxy appears to be a large, irregular spiral near to whose equatorial plane the Sun lies. Looking along this plane, the main bulk of the galactic stars are seen in projection as the Milky Way. The Galaxy also contains a great deal of the dark obscuring material which is, as we have seen, a feature of other spirals. This material betrays its presence by the sudden gaps or dark spaces occurring in the Milky Way and in the almost complete absence of extra-galactic nebulae in that region of the sky.

From a study of galactic Cepheids it is concluded that the length of the equatorial diameter of the Galaxy is at least 3×10^4 parsecs and that of its transverse diameter at the widest part is $2{\cdot}5 \times 10^4$ parsecs.[3] These dimensions, however, include widely scattered outlying clusters of stars: the main bulk of the galactic population lies near to the equatorial plane in a more restricted region of space.

There is definite evidence, obtained from the individual, or 'proper', motions of the so-called 'fixed' stars, that the Galaxy is rotating. The Sun, which is believed to lie at some $1{\cdot}6 \times 10^4$ parsecs from the galactic centre, is

moving as a result with a velocity of 300 km./sec. The period of rotation of the Galaxy is estimated at 10^8 years. Other spirals are believed to have comparable periods and it is therefore not surprising that their angular velocities have proved difficult to measure.

For a long time after the distances of the extra-galactic nebulae had been determined it was believed that they fell far short of the Galaxy in size. More accurate densitometer measurements [3] have recently suggested that this was a false conclusion due to too much emphasis having been laid on the brighter equatorial regions of the nebulae. The faint outlying clusters of stars, which are always included in measuring the Galaxy, would be barely visible at the distance of even the nearest spiral nebula. The revised dimensions of the Andromeda nebula are now given as $1{\cdot}94 \times 10^4$ and $1{\cdot}71 \times 10^4$ parsecs. But it is nevertheless true that small nebulae are exceedingly numerous. Diameters of 9000 parsecs (Messier 60) and 5500 parsecs (Large Magellanic Cloud) are not uncommon.

DISTRIBUTION OF NEBULAE IN SPACE

It is estimated that the number of nebulae visible in modern telescopes runs into millions. The surveys so far carried out reveal the existence of numerous clusters [3, 4] of nebulae some of which have hundreds of members. The Coma cluster contains 500 nebulae all within a sphere of radius $2{\cdot}5 \times 10^5$ parsecs. In a cluster, the nebulae must be within each other's gravitational attraction and so may be regarded as physically connected systems rather than as chance momentary collections. But, on the other hand, the distribution of nebulae taken as a whole presents every appearance of being a random one. Apparent magnitude being taken as a rough criterion of distance, there is a uniform large-scale distribution in which the clustering appears as a random unevenness.

A survey [5] of the distribution of nebulae recently (Jan. 1937) completed at the Mount Wilson Observatory has, as will be shown later, important bearings on cosmological theory. On some hundreds of photographs of representa-

tive regions of the sky, a count was made of the numbers of nebulae with apparent magnitudes not greater than each of the five successive limiting values

$$m = 18{\cdot}47,\ 19,\ 19{\cdot}4,\ 20,\ 21{\cdot}03.$$

The result of these observations is summarized in the formula

$$\log_{10} N = 0{\cdot}501m - 2{\cdot}758 \quad . \quad . \quad . \quad (1.5)$$

in which N is the number of nebulae over the whole sky of apparent magnitude $\leq m$. It is believed to be accurate to within the probable errors of observation which, however, are necessarily large. The chief source of uncertainty is in the limiting magnitudes, which lie near the threshold of visibility even with the 100-inch telescope. However, it appears unlikely that a more accurate formula will be available until the 200-inch telescope, now under construction in America, comes into use.

MASS OF A NEBULA

A study of the dynamics of the Virgo cluster of nebulae has led Sinclair Smith [6] to the conclusion that an average nebula of this cluster, which has several hundred members, has a mass equal to 2×10^{11} times that of the Sun. On the other hand, the dynamics of a single rotating nebula, together with the two or three observed angular velocities, indicate a mass of 10^9 times [1] that of the Sun. Since the Sun's mass is $1{\cdot}983 \times 10^{33}$ gr. this gives 2×10^{42} gr. for the mass of an average nebula. With this value of the mass, which is the one usually accepted, Shapley has calculated [7] that if all nebulae within distances corresponding to $m = +13$ were disintegrated and then evenly spread out in the corresponding volume of space, the average density of matter would be of the order of 10^{-30} gr./cm.^{3}

DISPLACEMENT OF THE SPECTRAL LINES

The most remarkable phenomenon presented by the extra-galactic nebulae is a displacement of their spectral lines towards the red end of the spectrum. This 'red-shift' is independent of wave-length and has all the char-

acteristics of a Doppler effect due to a velocity of recession.[8] It is a very well-marked and easily observed phenomenon. With the 100-inch telescope and a spectrograph of small dispersion (418 Å/mm. at $\lambda = 4500$), it is possible to photograph the spectra of nebulae as faint as the 18th apparent magnitude. Four lines H, K, $H\delta$ (λ 4101) and the G band (λ 4303–14) are usually clearly shown and determinations for over 170 nebulae are now available. Denoting by δ the displacement in wavelength λ, we have $\delta = \frac{d\lambda}{\lambda}$, whilst the corresponding Doppler velocity, v, is given by

$$v = c\delta$$

where c is the velocity of light. Plotting $\log_{10} v$ (v in km./sec.) against the apparent magnitude of the nebula observed, Hubble and Humason up to 1934 gave the relation[2, 9] between these quantities as

$$\log_{10} v = 0{\cdot}2m + 0{\cdot}507 \pm 0{\cdot}012 \quad . \quad . \quad (1.6)$$

This formula has now (1937) been revised owing to the accumulation of some further data. Unfortunately, the new formula contains the apparent magnitude *corrected for red-shift*. This correction is intended to allow for the diminution of brightness due to the displacement of the spectrum. But the correction-formula employed is a theoretical one whose use would be fully justified only if the nature of the universe and the curvature of space, which are the very problems we are trying to investigate, were already known. Hubble's new formula [10] is

$$\log_{10} v = 0{\cdot}2m_c + 0{\cdot}77 \quad . \quad . \quad . \quad (1.7)$$

where m_c is the corrected magnitude. In its general effect the correction diminishes m whilst leaving v unaltered. Hence part at least of the increase in the constant term in (1.7) as compared with (1.6) is compensation for the correction and is not due to the use of new observational material. We shall therefore not be far wrong in assuming that, in the relation between $\log_{10} v$ and *uncorrected apparent magnitude*, the constant term cannot be less than 0·51 nor greater than 0·77.

We also remark here that formulae (1.6) and (1.7) are deduced essentially for nebulae of apparent magnitudes between 10 and 17 which are presumably nearer to the Galaxy than the faint nebulae included in large numbers in the material from which the distribution formula (1.5) was derived.

Using (1.6) and (1.4) we obtain

$$\log_{10}(v/D) = 0{\cdot}2M - 0{\cdot}493 \quad . \quad . \quad (1.8)$$

But $\qquad M = 2{\cdot}5 \log E_0 - 2{\cdot}5 \log E.$

Hence $\qquad \log_{10}(v/D) = (constant) - 0{\cdot}5 \log_{10} E$

or

$$\frac{v}{D} = \frac{constant}{\sqrt{E}}.$$

There is therefore *a linear relation between velocity and distance*, *D*, since it seems fairly certain that the luminosity of all nebulae is statistically constant so that E is the same for all. The name *velocity-distance relation* is therefore given to the formula (1.6). Introducing $M = -14{\cdot}2$, $D = 10^6$ into (1.8), we obtain

$$v = 465 \text{ km./sec. per } 10^6 \text{ parsec} \quad . \quad . \quad (1.9)$$

On the other hand, using the maximum value, 0·77, of the constant term in (1.6) we obtain

$$v = 851 \text{ km./sec. per } 10^6 \text{ parsec} \quad . \quad (1.10)$$

The average velocity of recession of the nebulae may be regarded as lying within these limits.

Expressed in terms of the Doppler shift, δ, the velocity-distance formula (1.6) is

$$\log_{10}\delta = 0{\cdot}2m - 4{\cdot}967 \pm 0{\cdot}012 \quad . \quad . \quad (1.11)$$

or, using the maximum value of the constant term,

$$\log_{10}\delta = 0{\cdot}2m - 4{\cdot}707 \quad . \quad . \quad . \quad (1.12)$$

The rates of increase of δ corresponding to the velocities (1.9) and (1.10) are respectively

$$d\delta = 1{\cdot}6 \times 10^{-3}/10^6 \text{ parsec} \quad . \quad . \quad (1.13)$$

$$d\delta = 2{\cdot}8 \times 10^{-3}/10^6 \text{ parsec} \quad . \quad . \quad (1.14)$$

COSMOLOGICAL THEORY

With this brief survey of the observational material we turn to the theoretical aspects of the structure of the universe. We shall endeavour to show how a large-scale uniform distribution of nebulae is compatible with a velocity-distance relation and also how the observed distribution law can be used to analyse the character of space itself. Our first step, however, must be to acquaint ourselves with the tensor calculus and with generalized differential geometry.

BIBLIOGRAPHY

1. E. Hubble, *Astrophys. J.*, **79**, 8, 1934.
2. E. Hubble and M. L. Humason, *Astrophys. J.*, **74**, 43, 1931.
   E. Hubble, *Proc. Nat. Acad. Sci. Philad.*, **15**, 168, 1929.
   H. Shapley, *Proc. Nat. Acad. Sci. Philad.*, **22**, 8, 1936.
3. H. Shapley, *Mon. Not. Roy. Astr. Soc.*, **94**, 791, 1934.
4. H. Shapley, *Proc. Nat. Acad. Sci. Philad.*, **21**, 587, 1935.
5. E. Hubble, *Astrophys. J.*, **84**, 517, 1936.
6. S. Smith, *Astrophys. J.*, **83**, 23, 1936.
7. H. Shapley and A. Ames, *Ann. Harvard Coll. Obs.*, **88**, No. 2, 1932.
8. M. L. Humason, *Proc. Nat. Acad. Sci. Philad.*, **15**, 167, 1929.
   H. Shapley, *Proc. Nat. Acad. Sci. Philad.*, **15**, 565, 1929.
   M. L. Humason, *Astrophys. J.*, **74**, 35, 1931; and **83**, 10, 1936.
9. E. Hubble and M. L. Humason, *Proc. Nat. Acad. Sci. Philad.*, **20**, 264, 1934.
10. E. Hubble, *Astrophys. J.*, 84, 158 and 270, 1936.

CHAPTER II

THE TENSOR CALCULUS

RIEMANNIAN SPACE

IN the ordinary space of three dimensions each point can be specified by three co-ordinates such as the three Cartesian co-ordinates (x^1, x^2, x^3).* The distance between two neighbouring points (x^1, x^2, x^3) and $(x^1 + dx^1, x^2 + dx^2, x^3 + dx^3)$ is du where

$$du^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 \quad . \quad . \quad (2.1)$$

It is possible to generalize this idea of space in two ways. Firstly, we can suppose space to have n dimensions instead of three. This will mean that each point of space will require n co-ordinates $(x^1, x^2, \ldots x^n)$ to specify it completely. We denote these co-ordinates collectively by (x). Secondly, we can suppose that the distance, ds, between two *neighbouring* points (x) and $(x + dx)$ is given by

$$ds^2 = \sum_{p,\,q=1}^{n} g_{pq}(x)dx^p dx^q \quad . \quad . \quad . \quad (2.2)$$

where the g_{pq} are functions of the co-ordinates (x) and may therefore vary from point to point. These functions are symmetrical in their indices so that $g_{pq} = g_{qp}$, $(p, q = 1, 2, \ldots n)$.

A space of n dimensions in which the distance *between neighbouring points only* is defined *a priori* is called a Riemannian space and the expression (2.2) is called the metric. It will be noticed that ordinary three-dimen-

* The indices 1, 2, 3 serve to distinguish the three co-ordinates from one another. The square of x^2 will be written $(x^2)^2$.

sional space falls under the category of Riemannian space, since (2.1) is the special case of (2.2) in which

$$g_{11} = g_{22} = g_{33} = 1$$

and
$$g_{12} = g_{23} = g_{31} = 0.$$

The functions g_{pq}, called the coefficients of the metric, are not necessarily positive, but it is always assumed that the determinant

$$g = \begin{vmatrix} g_{11} & g_{12} \ldots & g_{1n} \\ g_{21} & g_{22} & g_{2n} \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \\ g_{n1} & g_{n2} & g_{nn} \end{vmatrix}$$

is never zero.

The quantity ds we shall call the *interval.*

TRANSFORMATIONS OF CO-ORDINATES

Fixing our attention on some point of a Riemannian space we may imagine that the co-ordinates by which the point is described are altered in some way. A simultaneous change of description of all the points of the space from co-ordinates (x) to co-ordinates (x') is called a transformation of co-ordinates and is defined by a set of n equations

$$x'^r = f^r(x^1, x^2, \ldots x^n) \qquad (r = 1, 2, \ldots n) \quad . \quad (2.3)$$

We impose the restriction on the functions f^r that they are soluble for the (x) in terms of the (x') so that

$$x^r = g^r(x'^1, x'^2, \ldots x'^n) \qquad (r = 1, 2, \ldots n) \quad . \quad (2.4)$$

If we take differentials of the equations (2.3) we obtain

$$dx'^r = \sum_{q=1}^{n} \frac{\partial f^r}{\partial x^q} dx^q = \sum_{q=1}^{n} \frac{\partial x'^r}{\partial x^q} dx^q \qquad (r = 1, 2, \ldots n) \quad . \quad (2.5)$$

SCALAR, VECTOR, INNER PRODUCT

A quantity $q(x)$ whose value, though perhaps not its formal expression, is unaltered by all permissible transformations of co-ordinates is called a *scalar.* Thus, in three-dimensional geometry, if we permit ourselves only those transformations corresponding to translating the

origin of co-ordinates and to rotating the axes about the origin, we know that du^2 in (2.1) remains invariant. Hence its value is unaltered and it is therefore a scalar.

Consider again n quantities $(V^1, V^2, \ldots V^n)$, each V^r being a known function of the (x), and let the transformation (2.3) be performed on the co-ordinates. Then the n quantities are said to be the components of a *contravariant vector* if they transform according to the same rule as do the differentials of the co-ordinates. Thus by (2.5), the V^r transform to V'^r where

$$V'^r = \sum_{q=1}^{n} \frac{\partial x'^r}{\partial x^q} V^q, \quad (r = 1, 2, \ldots n). \quad . \quad (2.6)$$

Incidentally we note that the differentials $(dx^1, dx^2, \ldots dx^n)$ must themselves be the components of a contravariant vector.

Another type of vector whose components are $(U_1, U_2, \ldots U_n)$ is called *covariant* if it transforms according to the law

$$U'_r = \sum_{q=1}^{n} \frac{\partial x^q}{\partial x'^r} U_q, \quad (r = 1, 2, \ldots n) \quad . \quad (2.7)$$

The relationship between covariant and contravariant vectors depends on the following theorem. *If* $(V^1, V^2, \ldots V^n)$ *and* $(U_1, U_2, \ldots U_n)$ *are any two vectors, one contravariant and the other covariant, then the sum of products,* $\sum_{q=1}^{n} V^q U_q$, *is always a scalar.* To prove this we form the corresponding sum in terms of the transformed components

$$\begin{aligned} \sum_{q=1}^{n} V'^q U'_q &= \sum_{q=1}^{n} \left(\sum_{l=1}^{n} \frac{\partial x'^q}{\partial x^l} V^l \right) \left(\sum_{m=1}^{n} \frac{\partial x^m}{\partial x'^q} U_m \right) \\ &= \sum_{m=1}^{n} \sum_{l=1}^{n} \frac{\partial x^m}{\partial x^l} V^l U_m. \end{aligned}$$

But the co-ordinates x^l, x^m are necessarily independent, hence

$$\frac{\partial x^m}{\partial x^l} = \begin{cases} 0 \text{ if } l \neq m \\ 1 \text{ if } l = m \end{cases} \quad . \quad . \quad . \quad . \quad (2.8)$$

From this it follows at once that

$$\sum_{q=1}^{n} V'^q U'_q = \sum_{l=1}^{n} V^l U_l$$

so that the value of the sum is unaltered by transformation and is therefore a scalar.

This sum of products of corresponding components of a covariant and a contravariant vector is called the *inner product* of the two vectors. The inner product of an arbitrary covariant with an arbitrary contravariant vector is therefore always a scalar.

SUMMATION CONVENTION

A noteworthy characteristic of the preceding formulae is that any term in which an index occurs twice, once in the 'upper', or contravariant, and once in the 'lower', or covariant, position, is invariably prefixed by a summation sign covering the values of the index from 1 to n. This is illustrated by the index 'q' in formulae (2.5), (2.6) and (2.7). We shall therefore in future omit the summation sign before such terms and understand that the double presence of an index in a term is to imply summation of that term over all values 1 to n of the index. Thus the expression for the metric will simply be

$$ds^2 = g_{pq}(x)dx^p dx^q$$

and (2.6), (2.7) will be written

$$V'^r = \frac{\partial x'^r}{\partial x^q} V^q, \qquad U'_r = \frac{\partial x^q}{\partial x'^r} U_q.$$

TENSORS

The definitions of scalars and vectors which we have given depend essentially on the laws of transformation of these entities. It is not difficult to extend the idea to entities with laws of transformation analogous to, but more complicated than, those of vectors. Consider for instance a set of n^2 symbols $T^{pq}(x)$ $(p, q = 1, 2, \ldots n)$ whose law of transformation is

$$T'^{pq} = \frac{\partial x'^p}{\partial x^a} \frac{\partial x'^q}{\partial x^b} T^{ab} \quad (p, q = 1, 2 \ldots n) \,. \qquad (2.9)$$

The complete set of components T^{pq} is said to form a *contravariant tensor of rank two.* Similarly the sets of n^2 symbols $T_{pq}\,(p, q = 1, 2, \ldots n)$ and $T^p_q\,(p, q = 1, 2, \ldots n)$, with the transformation laws

$$T'_{pq} = \frac{\partial x^a}{\partial x'^p}\frac{\partial x^b}{\partial x'^q}T_{ab}, \qquad T'^p_q = \frac{\partial x'^p}{\partial x^a}\frac{\partial x^b}{\partial x'^q}T^a_b, \qquad (2.10)$$

define respectively a *covariant tensor of rank two* and a *mixed tensor of rank two.* The *rank* of a tensor thus indicates only the *number* of its indices per component, not the covariant or contravariant character of these indices. Comparison of (2.9), (2.10) with (2.6), (2.7) will show that a contravariant index of a tensor is associated with a law of transformation similar to that of a contravariant vector, and that a covariant index follows the rule of transformation of a covariant vector.

Proceeding in this way, tensors of higher and higher rank are built up. For instance, the n^4 symbols B^p_{qrl} together form a tensor of rank four with three covariant indices and one contravariant. The law of transformation is here

$$B'^p_{qrl} = \frac{\partial x'^p}{\partial x^a}\frac{\partial x^b}{\partial x'^q}\frac{\partial x^c}{\partial x'^r}\frac{\partial x^d}{\partial x'^l}B^a_{bcd}.$$

In particular, vectors are themselves tensors, their rank being unity, whilst scalars may be regarded as tensors of rank zero.

We shall in future denote a vector or tensor by a single one of its components, so that, for instance, the tensor whose law of transformation is (2.9) we shall refer to simply as 'the tensor T^{pq}'.

CONTRACTION

When a mixed tensor has been defined there is an important method of producing from it a new tensor, of rank two lower, which is called 'contraction'. We select a covariant and a contravariant index of the given tensor and then sum all the components for which these indices are equal. The result is a component of the new tensor.

As illustrations, consider first the tensor of rank four,

B^p_{qrl}, and select the index 'p' and the last covariant index 'l'. The contracted tensor is then

$$G_{qr} = B^p_{qrp}.$$

To verify that this is a tensor of rank two, we note that

$$G'_{qr} = B'^p_{qrp} = \frac{\partial x'^p}{\partial x^a}\frac{\partial x^b}{\partial x'^q}\frac{\partial x^c}{\partial x'^r}\frac{\partial x^d}{\partial x'^p}B^a_{bcd}$$

by (2.8) $$= \frac{\partial x^b}{\partial x'^q}\frac{\partial x^c}{\partial x'^r}B^a_{bca} = \frac{\partial x^b}{\partial x'^q}\frac{\partial x^c}{\partial x'^r}G_{bc}.$$

Hence the symbols G_{bc} have the transformation law of the components of a covariant tensor of rank two.

Secondly, let us apply the operation of contraction to a mixed tensor of rank two, say, T^p_q. The result is $T = T^p_p$ and it is easily proved that T is a tensor of rank zero, or a scalar. Thus associated with a mixed tensor of rank two there is a unique scalar obtained by contraction.

THE QUOTIENT THEOREM

When we are given a set of symbols, it is possible to determine their tensor character by forming their inner products with arbitrary covariant or contravariant vectors. This process, called the 'quotient theorem', is stated thus:

'*Any set of symbols whose inner product with an arbitrary covariant (or contravariant) vector is a tensor, are themselves the components of a tensor.*'

To prove the theorem for a set of n^2 symbols $T(pq)$ characterized by two indices and whose inner product with an arbitrary covariant vector U_q is known to be a contravariant vector V^p, we consider the transformation of

$$V^p = U_q T(pq).$$

We have

$$U'_q T'(pq) = V'^p = \frac{\partial x'^p}{\partial x^r}V^r = \frac{\partial x'^p}{\partial x^r}\{U_q T(rq)\}\ . \quad (2.11)$$

Again, in (2.7), interchanging the primed and unprimed letters we obtain

$$U_q = \frac{\partial x'^t}{\partial x^q}U'_t.$$

Hence (2.11) becomes

$$U'_q T'(pq) = \frac{\partial x'^p}{\partial x^r}\left\{\frac{\partial x'^t}{\partial x^q} U'_t\, T(rq)\right\}$$

or

$$U'_q\left\{T'(pq) - \frac{\partial x'^p}{\partial x^r}\frac{\partial x'^q}{\partial x^t}\, T(rt)\right\} = 0.$$

Since the vector U_q is arbitrary we must have

$$T'(pq) = \frac{\partial x'^p}{\partial x^r}\frac{\partial x'^q}{\partial x^t} T(rt)$$

so that the symbols $T(pq)$ transform by the law of a contravariant tensor of rank two and are therefore the components of such a tensor.

A similar proof holds for a set of symbols characterized by more than two indices.

THE MIXED FUNDAMENTAL TENSOR

So far we have considered tensors in the abstract. We shall now find a tensor of rank two which arises from the existence of co-ordinate transformations (2.3). We prove the following theorem.

' *The n^2 symbols δ^p_q where*

$$\delta^p_q = \begin{cases} 0 \text{ if } p \neq q \\ 1 \text{ if } p = q \end{cases} \quad (p,\ q = 1,\ 2,\ .\ .\ n)\ .\ \ (2.12)$$

are the components of a mixed tensor of rank two whose components have the same value in every co-ordinate system.

This follows at once from (2.8); for if δ'^p_q is the transform of δ^p_q, then

$$\delta'^p_q = \frac{\partial x'^p}{\partial x^a}\frac{\partial x^b}{\partial x'^q}\delta^a_b$$

by (2.12)
$$= \frac{\partial x'^p}{\partial x^a}\frac{\partial x^a}{\partial x'^q}$$

by (2.8)
$$= \begin{cases} 0 \text{ if } p \neq q \\ 1 \text{ if } p = q. \end{cases}$$

Hence $\delta'^p_q = \delta^p_q \quad (p,\ q = 1,\ 2,\ .\ .\ .\ n).$

This tensor is called the *mixed fundamental tensor*.

THE TWO FUNDAMENTAL TENSORS OF RIEMANNIAN SPACE

Two more tensors of rank two are defined by the metric itself. We assume that the interval between two nearby points is an intrinsic property of the two points. It must therefore be independent in magnitude of the co-ordinate system used to describe the points and so must be a scalar. Suppose that on changing the co-ordinate system from (x) to (x'), the coefficients g_{pq} become g'_{pq}. We have

$$ds^2 = g'_{pq}dx'^p dx'^q = g_{pq}dx^p dx^q,$$

so that by (2.5)

$$g'_{pq}\frac{\partial x'^p}{\partial x^a}\frac{\partial x'^q}{\partial x^b}dx^a dx^b = g_{pq}dx^p dx^q.$$

Since the differentials of the co-ordinates are independent, the last equation can only be true if

$$g'_{pq}\frac{\partial x'^p}{\partial x^a}\frac{\partial x'^q}{\partial x^b} = g_{ab}.$$

Interchanging the primed and unprimed letters, we obtain

$$g'_{ab} = g_{pq}\frac{\partial x^p}{\partial x'^a}\frac{\partial x^q}{\partial x'^b},$$

which is precisely the law of transformation of a covariant tensor of rank two. The tensor g_{pq} constitutes the *fundamental covariant tensor of rank two.*

Secondly, we define the *fundamental contravariant tensor of rank two*, g^{pq}, as follows. Consider the symbols defined by

$$g^{pq} = \frac{\textit{co-factor of } g_{pq} \textit{ in } g}{g} \qquad . \quad . \quad (2.13)$$

By the law of multiplication of determinants we have

$$\begin{aligned} g^{pq}g_{pr} &= 0 \text{ if } r \neq q \\ &= 1 \text{ if } r = q, \end{aligned}$$

so that $$g^{pq}g_{pr} = \delta^q_r \quad (q, r = 1, 2, \ldots n) \qquad . \quad (2.14)$$

To prove that the symbols g^{pq} are indeed the components of a contravariant tensor of rank two, we form first the

inner product, $g_{pt}V^t = U_p$, of the tensor g_{pq} with an arbitrary contravariant vector V^p. U_p is therefore an arbitrary covariant vector. We then have

$$g^{pq}U_p = g^{pq}g_{pt}V^t = \delta^q_t V^t = V^q.$$

Hence the inner product of the g^{pq} with an arbitrary covariant vector is a contravariant vector. By the quotient theorem, the g^{pq} must be the components of a contravariant tensor of rank two, which is the *fundamental contravariant tensor* of the Riemannian space.

ASSOCIATED TENSORS

The fundamental tensors enable us to perform the operations of 'raising' and 'lowering' the indices of a tensor which change a covariant index into a contravariant one and vice versa. This process really consists of forming the inner product of a given tensor with one or other of the fundamental tensors. The same letter is used to denote the resulting new tensor. Applied to the vectors V^p and U_p the process gives

$$V_q = g_{pq}V^p \text{ and } U^q = g^{pq}U_p \quad (q = 1, 2, \ldots n),$$

respectively. Again, from a tensor of rank two, T^{pq}, we obtain

$$T^p_q = g_{qr}T^{pr}, \qquad T_{pq} = g_{pt}T^t_q = g_{pt}g_{qr}T^{tr}.$$

It is easily proved that the results of the operations represented by the right-hand sides of these equations do, in fact, give tensors of the types indicated. Consider, for instance, the law of transformation of V_q. We have

$$V'_q = g'_{pq}V'^p = \left(\frac{\partial x^a}{\partial x'^p}\frac{\partial x^b}{\partial x'^q}g_{ab}\right)\left(\frac{\partial x'^p}{\partial x^c}V^c\right) = \frac{\partial x^a}{\partial x_c}\frac{\partial x^b}{\partial x'^q}g_{ab}V^c$$

$$= \delta^a_c\frac{\partial x^b}{\partial x'^q}g_{ab}V^c = \frac{\partial x^b}{\partial x'^q}g_{bc}V^c = \frac{\partial x^b}{\partial x'^q}V_b,$$

which is precisely the law of transformation of a covariant vector.

Tensors obtained from one another by this process are said to be *associated tensors*.

It is interesting to note that in ordinary three-dimensional space whose metric is (2.1) there is no formal difference between associated covariant and contravariant tensors if Cartesian co-ordinates are used. Raising and lowering suffixes here reduces to multiplication by unity.

GEODESICS

Hitherto we have confined our attention to the changes which tensors undergo when the co-ordinates of a particular point in a Riemannian space are changed. We shall now consider the changes which occur when we keep to the *same co-ordinate system* but *proceed from point to point* within the space. And in order to be able to travel through the space we shall begin by defining certain curves, the geodesics, whose properties are analogous to those of straight lines in ordinary three-dimensional space.

Consider two points P_0 and P_1 in the Riemannian space and any curve joining them. Such a curve is defined by a set of equations

$$x^r = F^r(\mu) \quad (r = 1, 2, \ldots n) \quad . \quad . \quad (2.15)$$

where μ is some parameter varying from point to point of the curve. By substituting (2.15) into (2.2) and then integrating with respect to μ, we can express the interval s measured along the curve in terms of μ. The geodesic joining P_0 and P_1 is then defined as that curve for which the interval s between P_0 and P_1 has a stationary value compared with the interval measured along any other neighbouring path joining the two points. The differential equations of the geodesics can be found as follows, the finite equations of these curves being unattainable without a precise knowledge of the functions g_{pq}.

Suppose that (2.15) represents the finite equations of the geodesic joining P_0 and P_1, so that the interval s is

$$s = \int_{\mu_0}^{\mu_1} \sqrt{\left\{g_{pq}(x)\,\frac{dx^p}{d\mu}\,\frac{dx^q}{d\mu}\right\}}d\mu \quad . \quad . \quad (2.16)$$

where μ_0, μ_1 are the values of μ at P_0 and P_1 respectively.

Any curve joining P_0 and P_1, and always lying close to the geodesic, will have equations of the form

$$\bar{x}^r = x^r + \varepsilon\omega^r = F^r(\mu) + \varepsilon\omega^r(\mu), \quad . \quad (2.17)$$

where $\omega^r = 0$ at $\mu = \mu_0$ and $\mu = \mu_1$ and ε is a small quantity whose square and higher powers we neglect. If then $\bar{s}$ is the interval along the curve (2.17) between P_0 and P_1 we have

$$\bar{s} = \int_{\mu_0}^{\mu_1} \sqrt{\left\{g_{pq}(\bar{x})\,\frac{d\bar{x}^p}{d\mu}\,\frac{d\bar{x}^q}{d\mu}\right\}}d\mu.$$

Neglecting all powers of ε above the first, we obtain

$$\bar{s} - s = \int_{\mu_0}^{\mu_1} \sqrt{\left\{g_{pq}\frac{dx^p}{d\mu}\,\frac{dx^q}{d\mu} + \varepsilon\left(\frac{\partial g_{pq}}{\partial x^l}\,\frac{dx^p}{d\mu}\,\frac{dx^q}{d\mu}\,\omega^l + 2g_{pq}\frac{dx^q}{d\mu}\,\frac{d\omega^p}{d\mu}\right)\right\}}d\mu$$
$$- \int_{\mu_0}^{\mu_1} \sqrt{\left\{g_{pq}\,\frac{dx^p}{d\mu}\,\frac{dx^q}{d\mu}\right\}}d\mu,$$

and, remembering that $\sqrt{\{g_{pq}dx^p dx^q\}} = ds$, this can be written

$$\bar{s} - s = \frac{\varepsilon}{2}\int_{\mu_0}^{\mu_1}\left\{\left(\omega^l\frac{\partial g_{pq}}{\partial x^l}\,\frac{dx^p}{d\mu}\,\frac{dx^q}{d\mu} + 2g_{pq}\frac{dx^q}{d\mu}\,\frac{d\omega^p}{d\mu}\right)\frac{d\mu}{ds}\right\}d\mu$$

We can now simplify the calculation by assuming (if $s \neq 0$) that the parameter μ is identical with the interval s measured along the geodesic. For in this case we have $d\mu/ds = 1$ and

$$\bar{s} - s = \frac{\varepsilon}{2}\int_{s_0}^{s_1}\left\{\omega^l\frac{\partial g_{pq}}{\partial x^l}\,\frac{dx^p}{ds}\,\frac{dx^q}{ds} + 2g_{pq}\frac{dx^q}{ds}\,\frac{d\omega^p}{ds}\right\}ds,$$

the x^q, ω^p now being regarded as functions of s. If we integrate the second term in the last equation once by parts we obtain

$$\bar{s} - s = \frac{\varepsilon}{2}\int_{s_0}^{s_1}\left\{\frac{\partial g_{pq}}{\partial x^l}\,\frac{dx^p}{ds}\,\frac{dx^q}{ds} - 2\frac{d}{ds}\left(g_{lq}\frac{dx^q}{ds}\right)\right\}\omega^l ds$$
$$+ \varepsilon\left[g_{lq}\frac{dx^q}{ds}\omega^l\right]_{s_0}^{s_1}.$$

But the functions ω^l vanish at $s = s_0$ and $s = s_1$. Hence $\bar{s} - s$

$$= \frac{\varepsilon}{2}\int_{s_0}^{s_1} \omega^l \left\{ \frac{\partial g_{pq}}{\partial x^l}\frac{dx^p}{ds}\frac{dx^q}{ds} - 2g_{lq}\frac{d^2x^q}{ds^2} - 2\frac{\partial g_{lq}}{\partial x^p}\frac{dx^p}{ds}\frac{dx^q}{ds} \right\} ds \quad (2.18)$$

If therefore the interval s is to have a stationary value for the geodesic compared with neighbouring curves, $\bar{s} - s$ must be zero for all functions ω^l. This is possible only if the coefficient of each ω^l in the integrand of (2.18) is separately zero. We conclude that the differential equations of the geodesics are the n equations

$$g_{lq}\frac{d^2x^q}{ds^2} + \frac{1}{2}\left(\frac{\partial g_{lq}}{\partial x^p} + \frac{\partial g_{lp}}{\partial x^q} - \frac{\partial g_{pq}}{\partial x^l}\right)\frac{dx^p}{ds}\frac{dx^q}{ds} = 0, \quad (l = 1, 2, \ldots n).$$

If in these equations we multiply by g^{lk}, perform the summation indicated by the double index and use (2.14), we obtain

$$\frac{d^2x^k}{ds^2} + \{pq, k\}\frac{dx^p}{ds}\frac{dx^q}{ds} = 0, \quad (k = 1, 2, \ldots n), \quad (2.19)$$

where

$$\{pq, k\} = \tfrac{1}{2}g^{lk}\left(\frac{\partial g_{lq}}{\partial x^p} + \frac{\partial g_{lp}}{\partial x^q} - \frac{\partial g_{pq}}{\partial x^l}\right).$$

The equations (2.19) are the equations of the geodesics in their standard form. It is to be observed that they involve the fundamental tensors and their first derivatives only.

The symbols $\{pq, k\}$ are called the Christoffel symbols for the metric. They are not tensors, as may be verified on using the transformation laws for the tensors g^{pq}, g_{pq}. It is clear from their definition that

$$\{pq, k\} = \{qp, k\}.$$

The equations (2.19) obviously possess the integral

$$g_{pq}\frac{dx^p}{ds}\frac{dx^q}{ds} = 1 \quad . \quad . \quad . \quad . \quad . \quad (2.20)$$

obtained by dividing the expression (2.2) by ds throughout.

We have obtained the equations of the geodesics from the stationary property of the interval, a property which

is clearly independent of the particular co-ordinate system used. The equations (2.19) must therefore be invariant in form for changes of co-ordinates.

NULL-GEODESICS

A second type of geodesic is obtained from the assumption that the interval between any two points on the geodesic is zero. These curves, called null-geodesics, are therefore characterized by having the integral

$$h = g_{pq}\frac{dx^p}{d\mu}\frac{dx^q}{d\mu} = 0, \quad . \quad . \quad (2.21)$$

where μ is any non-zero parameter. The differential equations of the null-geodesics may be obtained as follows. If the points P_0 and P_1 are now regarded as lying on the null-geodesic, the integral

$$I = \int_{\mu_0}^{\mu_1} h \, d\mu$$

must be zero along the curve. Along a neighbouring curve with equations of the form (2.17), we have, neglecting powers of ε greater than the first,

$$\bar{I} = \int_{\mu_0}^{\mu_1} \bar{h}\, d\mu = \int_{\mu_0}^{\mu_1}\left\{g_{pq}(\bar{x})\frac{d\bar{x}^p}{d\mu}\frac{d\bar{x}^q}{d\mu}\right\}d\mu$$

$$= \int_{\mu_0}^{\mu_1}\left\{g_{pq}\frac{dx^p}{d\mu}\frac{dx^q}{d\mu} + \varepsilon\left(\frac{\partial g_{pq}}{\partial x^l}\frac{dx^p}{d\mu}\frac{dx^q}{d\mu}\omega^l + 2g_{pq}\frac{dx^q}{d\mu}\frac{d\omega^p}{d\mu}\right)\right\}d\mu.$$

The first term in the integrand vanishes by (2.21). Hence

$$\bar{I} = \varepsilon\int_{\mu_0}^{\mu_1}\omega^l\left\{\frac{\partial g_{pq}}{\partial x^l}\frac{dx^p}{d\mu}\frac{dx^q}{d\mu} - 2\frac{d}{d\mu}\left(g_{lq}\frac{dx^q}{d\mu}\right)\right\} + \varepsilon\left[g_{lq}\frac{dx^q}{d\mu}\omega^l\right]_{\mu_0}^{\mu_1}.$$

The integrated term vanishes as before. Hence if the property $h = 0$ holds to the first order for all curves in the neighbourhood of the null-geodesic joining P_0 and P_1, we must have $\bar{I} = I = 0$. From this it follows that

$$\frac{\partial g_{pq}}{\partial x^l}\frac{dx^p}{d\mu}\frac{dx^q}{d\mu} - 2\frac{d}{d\mu}\left(g_{lq}\frac{dx^q}{d\mu}\right) = 0, \quad (l = 1, 2, \ldots n).$$

As before these equations reduce to

$$\frac{d^2x^k}{d\mu^2} + \{pq, k\}\frac{dx^p}{d\mu}\frac{dx^q}{d\mu} = 0, \quad (k = 1, 2, \ldots n), \quad (2.22)$$

which are the standard forms of the equations of the null-geodesics.

GEODESICS IN THREE-DIMENSIONAL SPACE

In three-dimensional space with metric (2.1), all the Christoffel symbols are zero. The equations of the geodesics reduce to

$$\frac{d^2x^k}{ds^2} = 0, \quad (k = 1, 2, 3).$$

In this case the finite equations can also be found. They are the well-known equations of the straight line in Cartesian co-ordinates :

$$\frac{x^1 - a}{l} = \frac{x^2 - b}{m} = \frac{x^3 - c}{n} = s,$$

where a, b, c are the co-ordinates of a point on the line. The constants l, m, n are related by

$$l^2 + m^2 + n^2 = 1$$

which corresponds to the integral (2.20). The null-geodesics have finite equations

$$\frac{x^1 - a}{l} = \frac{x^2 - b}{m} = \frac{x^3 - c}{i\sqrt{l^2 + m^2}} = \mu,$$

where the sum of the squares of the direction-cosines is now zero by (2.21). The null-geodesics of ordinary three-dimensional space are therefore *imaginary straight lines.*

COVARIANT DERIVATIVE

Having constructed the set of paths in Riemannian space represented by the geodesics, we are in a position to investigate the change which occurs in a vector or tensor when we imagine it to be moved from one point of the space to another along one of these curves. It will

be sufficient for our purpose to study the change produced in a contravariant vector V^p.

Since ds is a scalar, we can regard $\dfrac{dx^p}{ds}$, $(p = 1 \ldots n)$, as being the components of a contravariant vector. Moreover, if this vector satisfies (2.20) we can look upon it as giving the direction of the tangent to a geodesic. Let (x) be a point on the geodesic, $(x + dx)$ a neighbouring point on this curve and ds the interval along the curve. Then we can *define* the vector $V^p(x + dx)$ to be the result of moving $V^p(x)$ from (x) to $(x + dx)$ by postulating that

$$\frac{d}{ds}\left(V_q \frac{dx^q}{ds}\right) = \frac{d}{ds}\left(g_{pq} V^p \frac{dx^q}{ds}\right)$$

shall be a scalar, whatever pair of nearby points on the geodesic we may be thinking of.

This definition gives rise as follows to a mixed tensor of rank two called the *covariant derivative* of V^p. We have, using (2.19),

$$\frac{d}{ds}\left(g_{pq} V^p \frac{dx^q}{ds}\right) = \left(g_{pl}\frac{\partial V^p}{\partial x^m} + V^p \frac{\partial g_{pl}}{\partial x^m} - \{ml, q\} g_{pq} V^p\right)\frac{dx^m}{ds}\frac{dx^l}{ds}.$$

Also $\qquad g_{pq}\{ml, q\} = \frac{1}{2}\left(\dfrac{\partial g_{mp}}{\partial x^l} + \dfrac{\partial g_{lp}}{\partial x^m} - \dfrac{\partial g_{ml}}{\partial x^p}\right).$

Hence

$$\frac{d}{ds}\left(g_{pq} V^p \frac{dx^q}{ds}\right) = \left(\frac{\partial V^p}{\partial x^m} + \{mq, p\} V^q\right)\left(g_{pl}\frac{dx^l}{ds}\right)\frac{dx^m}{ds}.$$

Now the left-hand side of this equation is a scalar, whilst on the right we have the contravariant vector $\dfrac{dx^m}{ds}$ and the covariant vector $g_{pl}\dfrac{dx^l}{ds}$ forming an inner product with the symbols $\dfrac{\partial V^p}{\partial x^m} + \{mq, p\} V^q$. By the quotient theorem, the latter must be the components of a tensor, contravariant as to the index p and covariant as to m. We write

$V_{,m}^{p}$ for a component of this tensor, using the comma to indicate that the tensor is a covariant derivative, so that

$$V_{,m}^{p} = \frac{\partial V^p}{\partial x^m} + \{mq, p\} V^q \quad . \quad . \qquad (2.23)$$

This second-rank tensor is therefore *the covariant derivative of the original vector* V^p.

We may also apply the same notion of invariance along a geodesic to the inner product of a covariant vector U_p and the tangent to the geodesic. The resulting covariant derivative is

$$U_{p,m} = \frac{\partial U_p}{\partial x^m} - \{pm, q\} U_q \quad . \quad . \qquad (2.24)$$

Since in ordinary three-dimensional space, using Cartesian co-ordinates, the Christoffel symbols are all zero, it follows that covariant derivatives now reduce to ordinary partial derivatives of the components of a vector with respect to the co-ordinates.

DIVERGENCE. EQUATION OF CONTINUITY

If V^p is any contravariant vector and $V_{,m}^{p}$ its covariant derivative, there is a unique scalar obtained from the latter tensor by contraction. This scalar is variously denoted by $V_{,p}^{p}$ or by *div* V^p and its expression is

$$V_{,p}^{p} = \frac{\partial V^p}{\partial x^p} + \{qp, p\} V^q.$$

But

$$\{qp, p\} = \tfrac{1}{2} g^{pl} \left(\frac{\partial g_{ql}}{\partial x^p} + \frac{\partial g_{pl}}{\partial x^q} - \frac{\partial g_{qp}}{\partial x^l} \right) = \tfrac{1}{2} g^{pl} \frac{\partial g_{pl}}{\partial x^q},$$

and by the law of differentiation of a determinant applied to g,

$$\frac{\partial g}{\partial x^q} = (\textit{co-factor of } g_{pl} \textit{ in } g) \times \frac{\partial g_{pl}}{\partial x^q} = g g^{pl} \frac{\partial g_{pl}}{\partial x^q}.$$

Hence
$$\{qp, p\} = \frac{1}{2g} \frac{\partial g}{\partial x^q}.$$

We may therefore re-write $V^p_{,p}$ in the form

$$V^p_{,p} = \frac{\partial V^p}{\partial x^p} + \frac{1}{2g}\frac{\partial g}{\partial x^p} V^p = \frac{1}{\sqrt{g}}\frac{\partial(\sqrt{g}V^p)}{\partial x^p}.$$

The expression

$$div\ V^p = \frac{1}{\sqrt{g}}\frac{\partial(\sqrt{g}V^p)}{\partial x^p} \quad . \quad . \quad . \quad (2.25)$$

is taken to be the standard form for the divergence of the contravariant vector V^p. It is noteworthy that although the divergence has been derived by using the covariant derivative, yet it does not involve the Christoffel symbols but only the determinant g of the coefficients of the metric.

Comparing again with ordinary three-dimensional space whose metric is (2.1), we have $\sqrt{g} = 1$ in this case.

Hence $$div\ V^p = \frac{\partial V^1}{\partial x^1} + \frac{\partial V^2}{\partial x^2} + \frac{\partial V^3}{\partial x^3}.$$

We thus arrive at the well-known divergence formula of the ordinary vector-calculus.

In classical mechanics it often happens that a vector with four components is required to satisfy an *equation of continuity*. This is illustrated by the density and momentum of a perfect fluid in motion or by the charge and current vector of a system of moving electrical charges. Both these vectors satisfy an equation of the form

$$\frac{\partial j^4}{\partial t} + \sum_{r=1}^{3} \frac{\partial j^r}{\partial x^r} = 0,$$

where t is the time and (x^1, x^2, x^3) are spatial co-ordinates. In the hydrodynamic interpretation j^4 is the density and (j^1, j^2, j^3) the components of momentum of the fluid ; in the electrical case they stand for the charge-density and the current respectively.

In tensor analysis, a contravariant vector, V^p, is said to obey an equation of continuity in the Riemannian space (2.2) if, at every point, the vector satisfies the equation

$$\frac{\partial(\sqrt{g}V^p)}{\partial x^p} = 0 \quad . \quad . \quad . \quad . \quad . \quad (2.26)$$

where g is the determinant of the coefficients of the metric.

RIEMANN-CHRISTOFFEL TENSOR

In the ordinary calculus, the application of the process of partial differentiation to the first partial derivative, $\frac{\partial y}{\partial x^m}$, of a function y will produce a second partial derivative of the general form $\frac{\partial^2 y}{\partial x^m \partial x^n}$. In tensor calculus covariant differentiation can also be repeated. Thus let $V,^p_m$ be the first covariant derivative of a contravariant vector V^p and let us assume that

$$\frac{d}{ds}\left(V,^p_m g_{pq} \frac{dx^m}{ds} \frac{dx^q}{ds}\right)$$

is a scalar, where, as before, the vector $\frac{dx^r}{ds}$ is the tangent to a geodesic. On working this expression out, we deduce that the second covariant derivative of V^p is a tensor of rank three $V,^p_{rl}$ where

$$V,^p_{rl} = \frac{\partial V,^p_r}{\partial x^l} + \{ml, p\}V,^m_r - \{rl, m\}V,^p_m \quad . \quad (2.27)$$

Unlike ordinary partial differentiation, the *order* in which the successive differentiations are performed is of great importance. We shall now calculate the difference $V,^p_{rl} - V,^p_{lr}$.

Using the formula (2.23) in (2.27) we obtain, after some calculation,

$$V,^p_{rl} - V,^p_{lr} =$$

$$V^q\left[\frac{\partial\{qr, p\}}{\partial x^l} - \frac{\partial\{ql, p\}}{\partial x^r} + \{ml, p\}\{qr, m\} - \{rm, p\}\{ql, m\}\right].$$

Now the left-hand side is the difference of two arbitrary mixed tensors of rank three and is therefore a tensor of the same rank and character. On the right-hand side, we have the inner product of an arbitrary contravariant vector V^q with the expression enclosed in square brackets.

By the quotient theorem, it follows that the latter must be a component of a tensor of rank four, contravariant as to one index and covariant as to the other three. We denote this tensor by B^p_{qlr}, where

$$B^p_{qlr} = \frac{\partial\{qr, p\}}{\partial x^l} - \frac{\partial\{ql, p\}}{\partial x^r} + \{ml, p\}\{qr, m\} - \{rm, p\}\{ql, m\}.$$

This tensor is called the *Riemann-Christoffel*, or *curvature, tensor*. The reason for the latter name is that in the theory of two-dimensional surfaces this tensor is proportional to the ordinary Gaussian curvature at each point of the surface. Its special importance lies in being the next most complicated tensor after the fundamental tensors themselves, which can be constructed out of these and their derivatives. It is therefore intrinsically associated with the geometry of the Riemannian space.

In relativity theory we shall be more directly concerned with the *contracted Riemann-Christoffel tensor*, G_{ql}, obtained by equating p and r in the expression for B^p_{qlr} and summing. We obtain

$$G_{ql} = \frac{\partial\{qp, p\}}{\partial x^l} - \frac{\partial\{ql, p\}}{\partial x^p} + \{ml, p\}\{qp, m\} - \{pm, p\}\{ql, m\}$$

from which also

$$G^q_l = g^{qt}G_{lt} \quad . \quad . \quad . \quad . \quad (2.28)$$

Contracting once again we obtain the *scalar curvature*

$$G = G^q_q = g^{qt}G_{qt} \quad . \quad . \quad . \quad . \quad (2.29)$$

Referring once more to the ordinary space with metric (2.1), it is clear that every component of the curvature tensor is now zero. Any space in which this happens is said to be *of zero curvature or 'flat'*.

ABSOLUTE PARALLELISM [1]

The covariant derivatives provide us with information as to the changes which tensors undergo when they are moved along geodesics. They give, however, no indication as to what changes take place if the path is not a geodesic. It follows that there is no unique vector at a point P which can be said to be *parallel* to a given vector

at some other point Q. For parallelism is equivalent to the possibility of moving one vector into coincidence with another and we have just seen that the result of any such motion may depend on the path taken between P and Q. An additional geometrical hypothesis is therefore necessary if we wish to introduce the concept of parallelism between vectors situated at points distant from one another in a Riemannian space.

We imagine that at every point of the space n vectors with covariant components ${}^{\alpha}h_p$, or contravariant ${}_{\alpha}h^p$, are set up. Here α $(= 1, 2, \ldots n)$ specifies the vector and p $(= 1, 2, \ldots n)$ the component of the vector α in some co-ordinate system which we have selected. These vectors are defined to be such that

$$\left.\begin{aligned} g_{pq} = {}^{\alpha}h_p\,{}^{\alpha}h_q, \qquad g^{pq} &= {}_{\alpha}h^p\,{}_{\alpha}h^q \\ \text{whence} \qquad {}_{\alpha}h^p\,{}^{\alpha}h_q &= \delta^p_q\,. \end{aligned}\right\} \quad . \quad . \quad (2.30)$$

It is also assumed that

$${}_{\alpha}h^p\,{}^{\beta}h_p = \delta^{\beta}_{\alpha}. \quad . \quad . \quad . \quad . \quad . \quad (2.31)$$

The vectors so defined are called the *fundamental vectors*. They are not, of course, unique, so that many different sets of fundamental vectors can be chosen in the same Riemannian space.

To every vector V^p we can now make correspond the n scalars

$${}^{\alpha}V = {}^{\alpha}h_p V^p \qquad (\alpha = 1, 2, \ldots n)$$

by taking the inner product of V^p with each one of the fundamental vectors in turn. We can then *define a vector $V^p(x)$ at the point (x) to be parallel to a vector $U^p(x')$ at any other point (x') if*

$${}^{\alpha}V(x) = {}^{\alpha}U(x'), \qquad (\alpha = 1, 2, \ldots n).$$

This means that the two vectors have equal corresponding scalars when referred to the fundamental vectors at (x) and (x') respectively.

We are thus able to speak of an *absolute* or *distant parallelism between vectors* in a Riemannian space *relative to some selected set of fundamental vectors.*

PATHS OF AN ABSOLUTE PARALLELISM

Suppose in particular that V^p is the vector $\dfrac{dx^p}{ds}$ representing the tangent to a curve which possesses the property that its tangent is parallel to itself along the curve. On proceeding from a point (x) to the nearby point $(x + dx)$ on such a curve the *scalar increments* of the components of the tangent-vector must be zero. This means that $\dfrac{d}{ds}\left({}^{\alpha}h_p \dfrac{dx^p}{ds}\right) = 0$, $(\alpha = 1, 2, \ldots n)$. From these we can deduce the differential equations of the curve in the form

$$ {}^{\alpha}h_p \frac{d^2x^p}{ds^2} + \frac{\partial {}^{\alpha}h_p}{\partial x^l}\frac{dx^l}{ds}\frac{dx^p}{ds} = 0. $$

On multiplication by ${}_{\alpha}h^q$ and the use of the third of equations (2.30), we obtain

$$ \frac{d^2x^q}{ds^2} + {}_{\alpha}h^q \frac{\partial {}^{\alpha}h_p}{\partial x^l}\frac{dx^l}{ds}\frac{dx^p}{ds} = 0. $$

Since the summation in the second term extends over both l and p we can write the equations in the symmetrical form

$$ \frac{d^2x^q}{ds^2} + \Lambda^q_{pl}\frac{dx^l}{ds}\frac{dx^p}{ds} = 0 \quad (q = 1, 2, \ldots n) \quad (2.32) $$

where
$$ \Lambda^q_{pl} = \tfrac{1}{2}{}_{\alpha}h^q\left(\frac{\partial {}^{\alpha}h_p}{\partial x^l} + \frac{\partial {}^{\alpha}h_l}{\partial x^p}\right). $$

The curves with these differential equations are called the *paths of the absolute parallelism* defined by the fundamental vectors ${}_{\alpha}h^p$, ${}^{\alpha}h_p$. *In general, these paths are not the same as the geodesics of the space.* Moreover, they are not usually the same for different selections of fundamental vectors. The equations of the paths, however, are invariant in form for co-ordinate transformations which also leave invariant the equations (2.30) and (2.31).

BIBLIOGRAPHY

1. *Differential Invariants of Generalized Spaces.* T. Y. Thomas, Camb. Univ. Press, 1934.

CHAPTER III

PRINCIPLES OF GENERAL RELATIVITY

THE theory of the expanding universe can hardly be understood without some idea of the fundamental assumptions which lie at the basis of the general theory of relativity. We shall therefore give a brief sketch of these ideas in so far as they are necessary for our purpose.

SPACE-TIME

It is a fact of common sense that every occurrence in the physical world takes place at some definite place and at some definite time. Mathematically we particularize this statement by saying that every such occurrence, technically called an '*event*', requires four numbers to specify it. These numbers we denote by t, x^1, x^2, x^3, or collectively by (t, x^r), and we call them *co-ordinates*. The first co-ordinate we suppose to be *in some sense* a measurement of *time* and the other three to be similarly related to measurements of distance so that they give the *spatial position* of the event. Let us put aside for the moment the question of the exact physical meanings of the co-ordinates and suppose that we have somehow assigned co-ordinates to all events belonging to a set on which we have fixed our attention. We also assume that this group of events may be represented by a set of points in a four-dimensional Riemannian space whose metric is

$$ds^2 = g_{44}dt^2 - \frac{1}{c^2} \sum_{p,\, q=1}^{3} g_{pq}dx^p dx^q \quad . \quad . \quad (3.1)$$

In this expression g_{44}, g_{pq} are functions of (t, x^r), c is a constant and ds an invariant.

The Riemannian space in which the events are mapped in this way is called *Space-Time*. Our special problem is to investigate how all the events constituting the universe can be represented in one and the same space-time.

We shall also assume that it is possible to define *a priori* a curve in space-time by equations of the form

$$t = f^4(\mu), \quad x^p = f^p(\mu) \quad (p = 1, 2, 3),$$

where μ is some parameter. Eliminating μ from the last three equations by means of the first, we obtain the equations of the curve in the form

$$x^p = h^p(t) \quad (p = 1, 2, 3) \quad . \quad . \quad . \quad . \quad (3.2)$$

Since t is a time-co-ordinate, the equations (3.2) are what would be called the *integrals of equations of motion* in mechanics. A curve in space-time is therefore in some sense a graph of the motion of a dynamical system.

TIME-CO-ORDINATES AND PROPER-TIME

We now return to the question of the physical meanings which can be attributed to the co-ordinates (t, x^r). Let us imagine an observer who is provided with a standard clock and a standard rigid measuring-rod. These instruments we regard as indefinable. Let us also assume that our observer takes as his fundamental axiom that *the interval between two events alone has physical significance* in terms of the measurements made by his clock and his rigid measuring-rod. If physical meanings can be discovered for the co-ordinates (t, x^r) of the two events, it is only as a consequence of this basic hypothesis.

As a matter of convention the observer decides that the interval s shall always have the dimensions of time. He bestows the name *proper-time* on the interval.

With these preliminaries, let us visualize the observer about to set up a co-ordinate system for all events in the universe. He will obviously give himself the simplest co-ordinates to start with and regard himself as 'at rest' in his co-ordinate system. Let his co-ordinates be (0, 0, 0, 0). Then at some later time in his history his co-ordinates will be $(t_1, 0, 0, 0)$ since only his time-co-

ordinate can vary. These two events also differ in physical time only which, by hypothesis, is a difference of s between the two events. Physical time is measured by the observer's clock which we may suppose to read $s = 0$ at the event $(0, 0, 0, 0)$ and $s = s_1$, at the event $(t_1, 0, 0, 0)$. We have

$$s_1 = \int_0^{s_1} ds = \int_0^{t_1} \sqrt{g_{44}(t, 0, 0, 0)}dt \quad . \quad (3 \cdot 3)$$

It is immediately clear that before the observer can express t_1 in terms of the number s_1 he must be able to evaluate the integral (3.3) and therefore must know the function $g_{44}(t, x^r)$. Let us therefore make the further assumption that our observer knows the explicit forms of g_{44}, g_{pq} in (3.1) as functions of the co-ordinates and that he also knows the value of the constant c. We leave aside for the moment the question of how he arrives at this information.

The observer can now express t_1 as a function of s_1, thus obtaining the co-ordinate-time lapse corresponding to the measured proper-time s_1. It is clear that $s_1 = t_1$ only if $g_{44}(t, 0, 0, 0) = 1$ for $0 \leq t \leq t_1$.

The equation (3.3) also brings out a characteristic of the co-ordinates used in general relativity, viz. the co-ordinate t need not have the physical *dimensions* of time. This follows at once from the fact that it is the combination $\sqrt{g_{44}}dt$ $(= ds)$ which has the dimensions of time so that t might be a pure number provided that $\sqrt{g_{44}}$ carried the 'dimensions'.

DISTANCE

The observer can regard the two ends of his measuring-rod as two events with the *same* value of t but with *different* spatial co-ordinates. Let one end be the event $(t_0, 0, 0, 0)$, the other $(t_0, dx^1, 0, 0)$. These events are separated by an interval ds where

$$ds = \frac{i}{c}\sqrt{g_{11}(t_0, 0, 0, 0)}dx^1.$$

If the expressions of type $\sqrt{g_{11}}dx^1$ are to have the dimensions of length, clearly c must be a velocity. Moreover, ds alone has physical significance. The observer therefore identifies the quantity $\frac{c}{i}ds$ with the length, dl, of his rigid scale. The co-ordinate increment dx^1 is therefore given by

$$dl = \sqrt{g_{11}(t_0, 0, 0, 0)}dx^1 \quad . \quad . \quad . \quad (3.4)$$

The definition of a finite value, X^1, of the spatial co-ordinate separating the events $(t_0, 0, 0, 0)$ and $(t_0, X^1, 0, 0)$, follows at once by integration. Suppose that the observer has a scale of sufficient length, l, then X^1 is defined by

$$l = \int_0^{X^1} \sqrt{g_{11}(t_0, x^1, 0, 0)}dx^1,$$

where, in the integrand, x^1 alone varies.

A very important consequence of this definition is that X^1 corresponds to a *fixed* value of l only if g_{11} does not involve t. Conversely, if g_{11} *does* involve t, two events which have fixed co-ordinates x^1 may yet change their distance apart. This has an important application in the theory of the universe.

More generally, if the ends of the rigid scale are regarded as the events $(t_0, 0)$ and (t_0, dx^r), the length of the scale is $dl = \{\sum_{p,q=1}^{3} g_{pq}dx^p dx^q\}^{\frac{1}{2}}$. Since the co-ordinates are independent it is not possible to integrate this expression to obtain the definitions of finite values of the co-ordinates. One way out of this difficulty is to define the spatial co-ordinates in terms of a series of distance measurements made in a definite order. Thus consider the events $(t_0, 0, 0, 0)$ and $(t_0, X^1, X^2, 0)$. We may define X^1, X^2 by means of the measurements l_1, l_2 where

$$l_1 = \int_0^{X^1} \sqrt{g_{11}(t_0, x^1, 0, 0)}dx^1,$$

$$l_2 = \int_0^{X^2} \sqrt{g_{22}(t_0, X^1, x^2, 0)}dx^2.$$

Here we have first measured from (o, o, o) to (X^1, o, o) and then from (X^1, o, o) to (X^1, X^2, o). It is not difficult to see that the resulting definitions of X^1, X^2 are not necessarily the same as those obtained by measuring first from (o, o, o) to (o, X^2, o) and then from (o, X^2, o) to (X^1, X^2, o). But so long as we apply the same convention as to order to the measurement of the co-ordinates of all events, no real ambiguity can arise.

CO-ORDINATES IN GENERAL

We now approach the more difficult question of the meanings which the observer attaches to the co-ordinates of an event which has a *different time* as well as *different spatial co-ordinates* from those which he himself possesses.

We again suppose that the observer constitutes the event (o, o, o, o) and that (t_1, X^1, o, o) is some other event. Let us also assume that the observer has a second standard clock, precisely similar to his own, which leaves him at the event (o, o, o, o), when it reads zero, and travels to the event (t_1, X^1, o, o) where it arrives reading s_1. This motion must be supposed to take place along some curve in space-time defined *a priori* and having equations of type (3.2). This knowledge, we have already remarked, is equivalent to knowing *a priori* the equations of motion of the clock. The observer then takes the reading s_1 on this moving clock to be the measure of the interval separating the two events in question, so that

$$s_1 = \int_0^{t_1} \sqrt{\left\{g_{44}(t, x^1, o, o) - \frac{g_{11}(t, x^1, o, o)}{c^2}\left(\frac{dx^1}{dt}\right)^2\right\}}dt \quad (3.5)$$

In the integrand, both x^1 and t vary, the former being given as a function of the latter by equations of type (3.2). Hence t_1 can be found as a function of s_1 and the value of X^1 then follows from the equations of motion of the clock.

As an example, let us postulate that the clock traces out the curve in space-time whose equations are

$$x^1 = vt, \; x^2 = o, \; x^3 = o,$$

where v is a constant. In the language of mechanics

we should say that the clock moved with constant velocity v in the co-ordinate system. Hence

$$X^1 = vt_1 \quad . \quad . \quad . \quad (3.6)$$

and (3.5) gives

$$s_1 = \int_0^{t_1} \sqrt{\left\{g_{44}(t, vt, 0, 0) - g_{11}(t, vt, 0, 0)\frac{v^2}{c^2}\right\}} dt \quad . \quad (3.7)$$

Evaluating this integral we obtain a relation giving t_1 in terms of s_1 and v. The velocity v, however, is not arbitrary since we can imagine that the observer has some means of measuring the distance of the event $(t_1, X^1, 0, 0)$ at that moment of his own history which constitutes the event $(t_1, 0, 0, 0)$. If l is this distance, we have

$$l = \int_0^{X^1} \sqrt{g_{11}(t_1, x^1, 0, 0)}dx^1 = \int_0^{vt_1} \sqrt{g_{11}(t_1, x^1, 0, 0)}dx^1 \quad (3.8)$$

where, of course, only x^1 now varies in the integrand. This integral gives a second relation between v, t_1 and the measured distance l, which, with the relation given by (3.7), yields the value of t_1 (and, of course, of v) in terms of the measured quantities s_1 and l. The equation (3.6) then gives X^1, and the co-ordinates of $(t_1, X^1, 0, 0)$ are thus completely determined.

An ideal method of finding t_1 without having recourse to a measurement of l, is to assume that the clock moves infinitely slowly so that v is always approximately zero. The equation (3.7) then becomes

$$s_1 = \int_0^{t_1} \sqrt{g_{44}(t, 0, 0, 0)}dt.$$

The lapse of proper-time on the moving clock is therefore equal to that on the observer's stationary clock between the events (0, 0, 0, 0) and $(t_1, 0, 0, 0)$. The measurement of l then serves to determine X^1 directly, since t_1 is now known.

It is not difficult to extend the foregoing results to the case when two events differ by more than one spatial co-ordinate as well as in their time co-ordinates.

CO-ORDINATES IN SPECIAL RELATIVITY

The procedure outlined above can be carried out completely in the space-time of special relativity. This theory is characterized by the use of a metric of the form

$$ds^2 = dt^2 - \frac{1}{c^2}\{(dx^1)^2 + (dx^2)^2 + (dx^3)^2\}$$

in which therefore the coefficients of the metric are $g_{44} = 1$, $g_{pq} = \delta_{pq}$. The formula (3.7) now becomes

$$s_1 = \int_0^{t_1} \sqrt{\left(1 - \frac{v^2}{c^2}\right)} dt = t_1 \sqrt{\left(1 - \frac{v^2}{c^2}\right)}$$

whilst (3.8) yields

$$l = X^1 = vt_1.$$

By elimination of v we obtain at once

$$t_1 = \frac{1}{c}\sqrt{(s_1{}^2c^2 + l^2)}, \qquad X^1 = l \quad . \quad . \quad (3.9)$$

for the co-ordinates assigned by the observer to the event $(t_1, X^1, 0, 0)$.

THE GEOMETRIZATION OF PHYSICS

It will be remembered that we have endowed our observer with a considerable amount of information whose origin is not immediately obvious. He has been supposed to know the functional forms of g_{44}, g_{pq}, the value of c and the laws of motion of certain moving clocks as represented by curves in space-time. We shall now show how this knowledge can be attained on the basis of a single axiom which we may call the principle of the geometrization of physics. Its enunciation is as follows:

'*A distribution of matter and radiation in any region of space-time and the Riemannian geometry appropriate to that region have the same qualitative and quantitative properties.*'

Firstly with regard to the qualitative properties. These are best illustrated by examples. Thus, for instance, if the distribution of matter and radiation exhibits symmetry about a line, the coefficients of the metric will also possess symmetry about that line. More precisely, the coefficients

will not involve a co-ordinate representing rotations round the line. Again, if the distribution does not change with the time, the coefficients will not involve the time-co-ordinate, and so on. In fact, considerations such as these are usually sufficient to fix which co-ordinates do, and which do not, enter into the expressions for the coefficients of the metric.

The possession of the same quantitative properties is ensured by rules which fix the exact relationship between the mechanical characteristics of matter and radiation, such as density, momentum, energy and the values of the coefficients of the metric. They are, of course, arbitrary in the same sense that Newton's laws of motion are arbitrary, viz. their justification lies in their consequences. The names given to these rules are *Einstein's gravitational equations* and the *principle of equivalence*, respectively.

EINSTEIN'S GRAVITATIONAL EQUATIONS

In classical hydrodynamics a distribution of matter of density ρ and velocity components (v^1, v^2, v^3) is completely described by the values of its *energy-tensor* whose components are defined by

$$T^{44} = \rho,\ T^{4r} = T^{r4} = \rho v^r \quad (r = 1, 2, 3);$$
$$T^{lm} = \rho v^l v^m \quad (l, m = 1, 2, 3). \quad . \quad (3.10)$$

If the matter consists of moving particles, the motion can be analysed into a mass-motion with velocity-components (u^1, u^2, u^3) together with the random motions of the constituent particles about this average motion. The random motions are grouped together and give rise to a *pressure* or *stress* term in the energy-tensor, which we denote by p^{lm} $(l, m = 1, 2, 3)$. The expressions (3.10) are now

$$T^{44} = \rho,\ T^{4r} = T^{r4} = \rho u^r \ (r = 1, 2, 3);$$
$$T^{lm} = p^{lm} + \rho u^l u^m \ (l, m = 1, 2, 3) \quad . \quad (3.11)$$

In particular if the matter consists of a perfect fluid at rest as a whole, the value of its energy-tensor is

$$T^{44} = \rho,\ T^{ll} = p,\ (l = 1, 2, 3) \quad . \quad . \quad (3.12)$$

all other components being zero. In this expression, ρ

is the density and p the hydrostatic pressure of the perfect fluid.

It is also known that a distribution of radiation is characterized by an electromagnetic energy-tensor of the same rank as T^{pq}. The components of the electromagnetic tensor may therefore be added to those of the material energy-tensor to give a composite tensor. The presence of radiation is, from this point of view, equivalent to the presence of matter.

We generalize the concept of an energy-tensor to four dimensions by substituting the velocity-vector

$$\left(\frac{dt}{ds}, \frac{dx^1}{ds}, \frac{dx^2}{ds}, \frac{dx^3}{ds}\right)$$

for the three-dimensional velocity used above. We obtain

$$T^{pq} = \rho_0 \frac{dx^p}{ds}\frac{dx^q}{ds} \quad (p, q = 1, 2, 3, 4) \quad (3.13)$$

where for the sake of uniformity we have written $t = x^4$. In this expression ρ_0 is called the *proper-density* of the matter. From this definition we conclude that T^{pq} is a contravariant tensor of rank two in four dimensions.

If therefore the geometry of any region of space-time depends on the state of matter contained therein, it follows that the coefficients of the metric must depend on this energy-tensor T^{pq}. We cannot here enter into the reasons which led Einstein to his particular choice of relationship between these two quantities. Suffice it to say that the identification he adopted, expressed in terms of the mixed tensor T^p_q, was given by the ten equations

$$-\kappa T^p_q = G^p_q - \tfrac{1}{2}(G - 2\Lambda)\delta^p_q, \quad (p, q = 1, 2, 3, 4) \quad (3.14)$$

where $\kappa = \dfrac{8\pi\gamma}{c^2}$, γ is the gravitational constant, Λ is the so-called cosmical constant and G^p_q, G are the components of the mixed Riemann-Christoffel tensor and the scalar curvature of the metric (3.1), respectively.

The equations (3.14) are Einstein's gravitational equations. They imply essentially that *any characteristic of the*

density, momentum and energy of matter and radiation in space-time will also be a characteristic of the curvature of space-time.

Since G^p_q and G are functions of the g^{pq}, g_{pq} [Equs. (2.28), (2.29)] the equations (3.14) are soluble for the coefficients of the metric if the values of T^{pq} are known. Returning to our observer who is setting up his co-ordinate system, Einstein's equations evidently enable him to find the coefficients of the metric but, at the same time, involve him in a vicious circle. For he cannot measure the energy-tensor of the matter in space-time unless he has a co-ordinate system in which to do it, and he cannot set up his co-ordinate system until he, in effect, knows the values of the energy-tensor. This is indeed a great weakness of the theory of relativity. Our observer can only surmount it by assigning to the energy-tensor *a priori* values, which appear plausible on general physical grounds, in terms of some abstract co-ordinate system (t, x^r). He then finds the coefficients of the metric from equations (3.14) and finally, by working back to the physical meanings of his abstract co-ordinates, compares the hypothetical physical situation which he has built up with that which he actually observes. Illogical as this procedure appears to be, it does not differ essentially from that adopted in other branches of mathematical physics, and it certainly works in practice.

From the definition (3.13) of the energy-tensor we obtain

$$T^p_p = g_{lp}T^{lp} = \rho_0\left\{g_{44}\left(\frac{dt}{ds}\right)^2 - \frac{1}{c^2}\sum_{l,m=1}^{3} g_{lm}\frac{dx^l}{ds}\frac{dx^m}{ds}\right\} = \rho_0,$$

and by contracting the equations (3.14),

$$-\kappa\rho_0 = G - \tfrac{1}{2}(G - 2\Lambda)4 = -G + 4\Lambda.$$

Hence $$\Lambda = \tfrac{1}{4}(G - \kappa\rho_0).$$

The space-time of special relativity is flat so that $G = 0$. It is also assumed that this space-time corresponds to a complete absence of both matter and radiation and that therefore $\rho_0 = 0$. The cosmical constant is therefore also zero in the space-time of special relativity.

PRINCIPLE OF EQUIVALENCE

Einstein's gravitational equations deal with the properties and motion of matter in the bulk. A further hypothesis is necessary in order to define *a priori* the *motion of individual particles of matter and of light-rays.*

We have seen that there exist in any Riemannian space certain curves, the geodesics, whose equations depend only on the coefficients of the metric and which are invariant in form for co-ordinate transformations. Remembering that in classical mechanics the path of a free particle is defined to be a straight line, *i.e.* a geodesic of three-dimensional space, the theory of relativity asserts that the path, or *world-line, of a freely moving particle is a geodesic of space-time.* This ' principle of equivalence ' as it is called therefore states that the equations of motion of a free particle are

$$\frac{d^2x^k}{ds^2} + \{mn, k\}\frac{dx^m}{ds}\frac{dx^n}{ds} = 0 \quad (m, n, k = 1, 2, 3, 4) \qquad (3{\cdot}15)$$

where we have again, for uniformity's sake, written x^4 for t. In these equations s is the proper-time measured by a clock travelling with the material particle. Incidentally our observer has now a definition of the curve in space-time traced out by one of his moving clocks, which has an invariant character not possessed by the rule $\frac{dx^1}{dt} = v$ previously employed.

The hypothesis made with regard to the *world-lines of light-rays* is that they are the *null-geodesics of space-time.* We are now in a position to give a physical meaning to the constant c. Suppose that a light-ray travels from the event (t, x^r) to the event $(t + dt, x^r + dx^r)$. Since the world-line is a null-geodesic, the proper-time ds between the two events *measured by a clock travelling with the light-ray*, is zero. Hence

$$g_{44}(t, x^r)dt^2 = \frac{1}{c^2}\sum_{l, m=1}^{3} g_{lm}(t, x^r)dx^l dx^m \quad . \quad (3.16)$$

Now *on a clock at rest in the position* (x^1, x^2, x^3) the proper-time which elapses between the events (t, x^r) and $(t + dt, x^r)$ is $ds_0 = \sqrt{g_{44}(t, x^r)}dt$, whilst the distance between the events (t, x^r) and $(t, x^r + dx^r)$ is $dl = \left\{ \sum_{l, m=1}^{3} g_{lm}(t, x^r)dx^l dx^m \right\}^{\frac{1}{2}}$.

Hence by (3.16) we have

$$\frac{dl}{ds_0} = \pm c \quad . \quad . \quad . \quad . \quad . \quad (3.17)$$

The velocity dl/ds_0 is called the *velocity of light* at the event (t, x^r) and its value is equal to the universal constant c. Thus the null-geodesic hypothesis for the motion of light is consistent with the statement that the velocity of light in terms of measurements made with clocks and rigid scales has the same value at all points of space-time.

PRINCIPLE OF COVARIANCE

It will have been noticed that our expression for the metric of space-time, the Einstein gravitational equations and the equations of the geodesics are all tensor equations. This means that if we were to change the co-ordinates in terms of which ds^2 is expressed, the rules of the tensor calculus would at once give us the new forms of the Einstein equations and of the geodesics. Unfortunately, the tensor calculus will not give us the corresponding new meanings of the co-ordinates in terms of physical measurements. For these depend on integral relations such as (3.7) and (3.8), a type of equation to which tensor analysis does not apply. Nevertheless the *principle of covariance*, viz., '*The laws of nature are expressible in tensor form*', is clearly of great importance. It puts on an equal footing all abstract co-ordinate systems whilst still permitting us to prefer one co-ordinate system to another from the observational point of view.

THEORY OF GRAVITATION

The foregoing principles, abstract and arbitrary as they may appear, have given rise to a theory of gravitation

more nearly in accordance with observation than is Newton's. Moreover, the tensor character of all the equations enables us to enunciate this theory of gravitation in any co-ordinate system and in any space-time. In the following chapter we shall investigate the extent to which these same principles can be used to give a theory of the universe as a whole.

BIBLIOGRAPHY

E. A. Milne, *Science Progress*, No. 115, 534, 1935.
H. Dingle, *Philosophy*, **11**, No. 41, 1936.

CHAPTER IV

THE EXPANDING UNIVERSE

METRIC OF THE UNIVERSE

WE shall idealize the problem of the structure of the universe in the following way. Instead of the actual system of the spiral nebulae, differing from one another in size and shape and having a somewhat irregular distribution in space, we shall consider a system of similar material particles forming a perfect fluid evenly spread out in space. More precisely, in terms of some abstract co-ordinate system (t, x), the state of the fluid at any event (t, x^r) is to be identical with its state at any other event (t, x'^r) which has the *same value of t* as the first. Each of the material particles will, of course, trace out a geodesic in space-time.

The universe being thus filled with an isotropic and homogeneous fluid, we now require the metric corresponding to such a distribution of matter. We take the origin of spatial co-ordinates (x^1, x^2, x^3) to be at any one of the particles of the fluid. The latter must present spherical symmetry around the particle. The coefficients of the metric must therefore also present spherical symmetry round the origin, by the principle of the geometrization of physics. The metric is therefore of the form

$$ds^2 = g_{44}(t, r)dt^2 - \frac{1}{c^2} \sum_{p,\, q=1}^{3} g_{pq}(t, r)dx^p dx^q \qquad (4.1)$$

where $r^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$.

Bearing in mind the phenomenon of the recession of the nebulae, we assume our fluid to be in motion so that its

state alters with the time. We have, however, seen that distance changing with time is not incompatible with fixed position in the co-ordinate system [Equ. (3.4)]. This statement is, in fact, merely another way of saying that co-ordinate systems moving with certain specified points may be used. We now show that this idea enables us to simplify the metric (4.1).

Let us assume that the particles of our fluid have each got fixed co-ordinates (x^1, x^2, x^3) in the system we are employing and that the co-ordinate t measures proper-time on similar clocks attached to each of the particles. Since the world-lines of our particles are geodesics, the foregoing hypothesis is equivalent to the assumption that the integrals of the equations of a certain set of geodesics of the metric (4.1) are

$$t = s\,;\; x^1 = \textit{constant}\,;\; x^2 = \textit{constant}\,;\; x^3 = \textit{constant}. \quad (4.2)$$

Hence along these geodesics

$$\frac{dt}{ds} = 1\,;\quad \frac{dx^p}{ds} = 0 \quad (p = 1, 2, 3)$$

and
$$\frac{d^2t}{ds^2} = 0\,;\quad \frac{d^2x^p}{ds^2} = 0 \quad (p = 1, 2, 3).$$

Using these in the equations (2.19), we obtain

$$\{44, 1\} = \{44, 2\} = \{44, 3\} = \{44, 4\} = 0;$$

where x^4, as usual, stands for t. If q stands for any one of the numbers 1, 2, 3 or 4, these equations are

$$\tfrac{1}{2}g^{pq}\left\{2\frac{\partial g_{4p}}{\partial x^4} - \frac{\partial g_{44}}{\partial x^p}\right\} = 0$$

or
$$\delta_q^4 g^{44}\frac{\partial g_{44}}{\partial x^4} - \tfrac{1}{2}g^{qp}\frac{\partial g_{44}}{\partial x^p} = 0 \quad (q = 1, 2, 3, 4).$$

Putting $q = 4$, it follows that g_{44} does not involve x^4 *i.e.* t. Again the three equations obtained by putting q equal to 1, 2 and 3 respectively, are

$$g^{11}\frac{\partial g_{44}}{\partial x^1} + g^{12}\frac{\partial g_{44}}{\partial x^2} + g^{13}\frac{\partial g_{44}}{\partial x^3} = 0$$

$$g^{21}\frac{\partial g_{44}}{\partial x^1} + g^{22}\frac{\partial g_{44}}{\partial x^2} + g^{23}\frac{\partial g_{44}}{\partial x^3} = 0$$

$$g^{31}\frac{\partial g_{44}}{\partial x^1} + g^{32}\frac{\partial g_{44}}{\partial x^2} + g^{33}\frac{\partial g_{44}}{\partial x^3} = 0.$$

Since the determinant of the g^{pq} is not to be zero, we must have

$$\frac{\partial g_{44}}{\partial x^1} = \frac{\partial g_{44}}{\partial x^2} = \frac{\partial g_{44}}{\partial x^3} = 0.$$

The final conclusion is that g_{44} in (4.1) must be a constant whose value may be taken to be unity. We emphasize once again that t is now *proper-time* measured on the clock attached to any one of the particles.

It remains to find the coefficients g_{pq} $(p,\ q = 1,\ 2,\ 3)$. Writing the metric as

$$ds^2 = dt^2 - \frac{1}{c^2}\sum_{p,q=1}^{3} g_{pq}(t, r)dx^p dx^q,$$

we note that the isotropy and homogeneity of the fluid in space for a given value of t, must be reflected in the geometry of space. This is achieved if, for a given value of t, every point of space is interchangeable with every other. Such spaces have long been known from purely geometric theory (Schur's theorem),[1] their metrics having the form

$$du^2 = R^2(t)\frac{(dx^1)^2 + (dx^2)^2 + (dx^3)^2}{(1 + kr^2/4)^2}$$

where $r^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$. In this expression $R(t)$ is a real function of t and k is a constant such that k/R^2 is the scalar curvature of the three-dimensional space at the moment t. The value of k may be 1, 0 or -1.

We conclude that a universe full of isotropic and homogeneous matter will have a metric of the form

$$ds^2 = dt^2 - \frac{R^2(t)}{c^2}\frac{(dx^1)^2 + (dx^2)^2 + (dx^3)^2}{(1 + kr^2/4)^2}. \quad (4.3)$$

where t is the *proper-time* on any clock carried by a particle with *fixed* co-ordinates (x^1, x^2, x^3).

In the expression (4.3) the physical dimensions may be assigned as follows. The co-ordinate t has the dimensions of s which are therefore those of time. The function $R(t)$ is given the dimensions of length so that the co-ordinates x^1, x^2, x^3 are pure numbers and the dimensions of c must be those of velocity.

The spatial co-ordinates (x^1, x^2, x^3) may take on any values between $+\infty$ and $-\infty$ and the function $R(t)$ may be called the *radius of curvature* of space.

DENSITY AND PRESSURE

Using Einstein's gravitational equations (3.14) and calculating the values of G^p_q, G for the metric (4.3) we obtain the following non-zero components for the energy-tensor:

$$\kappa T_4{}^4 = -\Lambda + 3\left(\frac{R'}{R}\right)^2 + \frac{3kc^2}{R^2},$$

$$\kappa T_1{}^1 = \kappa T_2{}^2 = \kappa T_3{}^3 = -\Lambda + \left(\frac{R'}{R}\right)^2 + \frac{2R''}{R} + \frac{kc^2}{R^2},$$

where a prime denotes differentiation with respect to t. We define the hydrostatic pressure, p, of the fluid by

$$c^2 p = -T_1{}^1 = -T_2{}^2 = -T_3{}^3$$

and its average density, ρ, by $c^2\rho = T_4{}^4$.

Hence
$$8\pi\gamma p = \Lambda - \left(\frac{R'}{R}\right)^2 - \frac{2R''}{R} - \frac{kc^2}{R^2} \quad . \quad . \quad (4.4)$$

$$8\pi\gamma\rho = -\Lambda + 3\left(\frac{R'}{R}\right)^2 + \frac{3kc^2}{R^2} \quad . \quad . \quad . \quad (4.5)$$

where ρ, p are expressed in gm./cm.^{3} By analogy with (3.12) this distribution of matter is called a perfect fluid. It will be noticed that both p and ρ depend only on the time, and not on position, in accordance with the assumption of isotropy and homogeneity.

If the equation (4.5) is differentiated with respect to t we obtain

$$8\pi\gamma\rho' = \frac{6R'}{R}\left\{\frac{R''}{R} - \left(\frac{R'}{R}\right)^2 - \frac{kc^2}{R^2}\right\}.$$

Hence
$$\rho' = -\frac{3R'}{R}(\rho + p)$$

and finally
$$d(R^3\rho) + pdR^3 = 0 \quad . \quad . \quad . \quad . \quad (4.6)$$
This gives a relation between the increments of $R^3\rho$, R^3 and the pressure p.

RED-SHIFT OF SPECTRAL LINES

In the last two paragraphs we have built up a 'hypothetical' physical universe in the way suggested in Chapter III. We have now to investigate the properties of this universe and compare them with the observed properties of the actual universe. We have still a good deal of latitude left as the function $R(t)$ and the constant k are, as yet, arbitrary.

The crucial question is whether or not we can deduce from the metric (4.3) the result that the light emitted by one particle, when viewed from any other particle, will exhibit a red-shift of the spectral lines proportionately to the distance between the particles.

We first transform the metric (4.3) to polar coordinates (r, θ, φ) by

$$x^1 = r \sin\theta \cos\varphi,\ x^2 = r \sin\theta \sin\varphi,\ x^3 = r\cos\theta \quad (4.7)$$

and obtain

$$ds^2 = dt^2 - \frac{R^2(t)}{c^2}\frac{dr^2 + r^2d\theta^2 + r^2\sin^2\theta d\varphi^2}{(1 + kr^2/4)^2} \quad . \quad (4.8)$$

Our particles will now have fixed co-ordinates (r, θ, φ). Let P_0 be the particle at the origin of space co-ordinates and let P_1 be any other particle. We may without loss of generality suppose that the spatial co-ordinates of P_1 are $(r_1, 0, 0)$. Let us also assume that P_1 emits a train of light-waves, successive 'crests' leaving that particle at

times t_1 and $t_1 + dt_1$ as measured by the clock at P_1. Let the successive wave-crests reach P_0 at times t_0 and $t_0 + dt_0$, according to the clock at P_0. The light-waves have travelled along the null-geodesic joining P_1 and P_0. Hence by (4.8)

$$\int_{r_1}^{0} \frac{dr}{1 + kr^2/4} = -c\int_{t_1}^{t_0} \frac{dt}{R(t)} = -c\int_{t_1+dt_1}^{t_0+dt_0} \frac{dt}{R(t)}, \quad . \quad (4.9)$$

since along this null-geodesic θ, φ are constants and ds is always zero.

We assume for the moment that the second two integrals (4.9) can be evaluated so that r_1 has a definite meaning. If then dt_0, dt_1 are small quantities we obtain

$$\frac{dt_1}{R(t_1)} = \frac{dt_0}{R(t_0)} \quad . \quad . \quad . \quad . \quad (4.10)$$

It has already been shown [Equ. (3.17)] that the velocity of light at any event in space-time is c. Hence if λ is the wave-length of the light when it left P_1 and $\lambda + d\lambda$, the wave-length when it reached P_0, we have

$$\lambda = cdt_1, \quad \lambda + d\lambda = cdt_0 \quad . \quad . \quad . \quad (4.11)$$

the co-ordinate t measuring proper-time at both P_1 *and* P_0. The three equations (4.10) and (4.11) then yield at once for the Doppler shift

$$\delta = \frac{d\lambda}{\lambda} = \frac{R(t_0)}{R(t_1)} - 1 \quad . \quad . \quad . \quad (4.12)$$

It is clear that the light is *always reddened* as it arrives at P_0 provided that $R(t_0) > R(t_1)$. Since $t_0 > t_1$, this means that the *function* $R(t)$ *must increase with* t. It must be pointed out that relativity theory gives no reason why this should be so, the metric (4.3) being equally suitable for the description of a *contracting* universe in which all shifts would be towards the *violet*.

We can calculate the rate of change of δ with distance in the following way. Consider a second particle, P_1', whose spatial co-ordinates are $(r_1 + dr_1, 0, 0)$, emitting

light-waves which also reach P_0 at time t_0. Then, by definition, the distance from P_1 to P_1' is

$$dl = \frac{R(t_1)dr_1}{(1 + kr_1^2/4)} \quad . \quad . \quad . \quad . \quad (4.13)$$

and the velocity of light at the particle P_1 at time t_1 is

$$\pm c = \frac{dl}{dt}.$$

Hence the light must leave P_1' at the instant $t_1 - \frac{dl}{c}$ in order to arrive at P_0 at the instant t_0. If therefore δ' denotes the Doppler shift in the light from P_1', then

$$\delta' = \frac{R(t_0)}{R\left(t_1 - \frac{dl}{c}\right)} - 1 = \delta + \frac{R(t_0)R'(t_1)}{R^2(t_1)} \frac{dl}{c}$$

and so

$$\frac{d\delta}{dl} = \frac{1}{c}\frac{R'(t_1)}{R(t_1)}\cdot\frac{R(t_0)}{R(t_1)} \quad . \quad . \quad . \quad . \quad . \quad (4.14)$$

In this expression $d\delta/dl$ is the rate of change in the red-shift with the distance of the source of light, *this distance being measured at the moment of emission*; t_0 is the fixed instant of observation and t_1, the time of emission of the light which, of course, varies with the distance of the emitting particle.

The law (4.14) corresponds to a *proportionality of red-shift with distance* only if the time of travel of the light, $t_0 - t_1$, from all observed sources, is relatively so short that $R'(t_1)/R^2(t_1)$ is approximately constant. In applying the formula to the actual universe, this is assumed to hold and (4.14) becomes

$$\frac{d\delta}{dl} = \frac{1}{c}\frac{R'(t_0)}{R(t_0)} \quad . \quad . \quad . \quad . \quad . \quad (4.15)$$

by putting $t_1 = t_0$.

In future we shall denote $R(t_0)$, $R(t_1)$ by R_0, R_1 respectively and use a similar notation for the derivatives of $R(t)$.

COMPARISON WITH OBSERVATION

We take for the observed value of $d\delta$ the rough mean of the values (1.13) and (1.14) so that

$$d\delta = 2{\cdot}2 \times 10^{-3},\ dl = 10^{6} \text{ parsec} = 3{\cdot}1 \times 10^{24} \text{ cm.}$$

Using these values in (4.15) we obtain, in round numbers,

$$\frac{R_0'}{R_0} = 2 \times 10^{-17} \text{ sec.}^{-1} \quad . \quad . \quad . \quad (4.16)$$

for the rate of increase of R at the present day.

To examine the validity of the approximation by which (4.15) is obtained from (4.14), we note that distances of the order of 5×10^{7} parsecs are greater than the distances of the faintest nebulae from which the velocity-distance relation is established. Corresponding to such distances we have

$$\varepsilon = t_0 - t_1 = (5 \times 10^{7})(3{\cdot}26)(3 \times 10^{7})$$
$$= 5 \times 10^{15} \text{ sec.} \quad . \quad (4.17)$$

since 1 parsec $= 3{\cdot}26$ light-years and, approximately, 3×10^{7} sec. $= 1$ sidereal year. But neglecting squares and higher powers of ε we can reduce formula (4.14) to

$$\frac{d\delta}{dl} = \frac{1}{c}\frac{R_0'}{R_0} - \frac{\varepsilon}{c}\left\{\frac{R_0''}{R_0} - 2\left(\frac{R_0'}{R_0}\right)^2\right\}.$$

It follows that the approximation (4.15) is still just valid at 5×10^{7} parsecs provided that the acceleration, R_0''/R_0, is of the order of $(R_0'/R_0)^2$.

We conclude that a space-time of metric (4.3) accounts for the observed phenomenon of the recession of the nebulae provided that $R(t)$ is such that R_0'/R_0 has an approximately constant value of 2×10^{-17} sec.$^{-1}$ over the distances so far surveyed.

RADIUS OF THE VISIBLE UNIVERSE

We now seek further tests by which the possibility of using the space-time with metric (4.3) for the actual universe could be demonstrated. One such test is pro-

vided by the formula (4.9). Evaluating the first integral we obtain

$$c\int_{t_1}^{t_0}\frac{dt}{R(t)} = \frac{2}{\sqrt{k}}\tan\frac{\sqrt{k}r_1}{2}, \quad r_1, \quad \frac{2}{\sqrt{k}}\tanh\frac{\sqrt{k}r_1}{2} \qquad (4.18)$$

according as $k > 0$, $k = 0$ or $k < 0$. Now it might happen that $R(t)$ was of such a mathematical form that $\int_{-\infty}^{t_0}\frac{dt}{R(t)}$ was a finite number $S(t_0)$ depending on t_0. The equations (4.18) would then imply that, for observations made at the origin at the instant t_0, there was a maximum co-ordinate-distance, $a(t_0)$, beyond which no objects were visible. The value of $a(t_0)$ in the three cases would be

$$a(t_0) = \frac{2}{\sqrt{k}}\tan^{-1}\left\{\frac{c\sqrt{k}S(t_0)}{2}\right\}, \quad cS(t_0),$$

$$\frac{2}{\sqrt{k}}\tanh^{-1}\left\{\frac{c\sqrt{k}S(t_0)}{2}\right\}.$$

This possibility could, in principle, be checked observationally. Let us consider the case $k = 0$ for simplicity. The nebula with the greatest radial co-ordinate, visible at time t from the origin, is situated at $(a(t_0), \theta, \varphi)$ where

$$a(t_0) = cS(t_0).$$

At the later time $(t_0 + dt)$ this most distant nebula is the one situated at $a(t_0 + dt)$ where

$$a(t_0 + dt) = cS(t_0 + dt).$$

Using Taylor's theorem and the definition of $S(t_0)$ we obtain

$$\frac{da}{dt} = \frac{c}{R_0}.$$

This clearly gives a measure of the rate at which nebulae, previously invisible to an observer at the origin, swept into his field of vision. Therefore observations directed to discovering whether or not faint nebulae gradually appeared in positions previously empty would not only give us information as to the character of the function $R(t)$ for

the actual universe but would provide one of the most powerful arguments in favour of the present theory as against its rivals.

If the function $S(t)$ exists the corresponding value of $a(t)$ may be called the *radius of the visible universe* according to the observer at the origin.

THE VOLUME OF SPACE AND THE CONSTANT k

At any instant t the volume enclosed within radius r is, by (4.8),

$$V = R^3 \int_0^r \int_0^\pi \int_0^{2\pi} \frac{r^2 dr d\theta \sin\theta d\varphi}{(1 + kr^2/4)^3}$$

$$= 4\pi R^3 \int_0^r \frac{r^2 dr}{(1 + kr^2/4)^3}.$$

Considering first the case $k = +1$ we obtain

$$V_+ = 4\pi R^3 \Big[z - \tfrac{1}{4} \sin 4z \Big]_0^{\tan^{-1}\frac{1}{2}r}$$

where $$\tan z = \tfrac{1}{2}r.$$

Since the maximum value of r is $+\infty$, it follows that the total volume of space is *finite* at every instant and is of amount

$$V_+ = 2\pi^2 R^3.$$

Such a space is called *spherical* or *closed* and is said to have a *positive curvature*. Physically it has the property of containing only a finite number of cubic miles at any particular instant.

If, secondly, $k = 0$ then the volume enclosed within radius r is

$$V_0 = \frac{4\pi}{3} R^3 r^3.$$

For a given value of R, there are an *infinite* number of cubic miles in space and V_0 is proportional to $\frac{4\pi}{3} r^3$ as in ordinary three-dimensional geometry. Space in such a case is said to be *flat* or of *zero curvature*. It does not, however, follow that a *flat space* implies a *flat space-time*.

The dimensions of space-time being greater by unity than those of space, curvature of the former is compatible with flatness of the latter. The analogy in two dimensions is that of ruled surfaces, on which, in spite of their curvature, one-dimensional straight lines can be drawn.

Lastly, suppose $k = -1$. Then

$$V_{-} = 4\pi R^3 \Big[\tfrac{1}{4} \sinh 4z - z \Big]_0^{\tanh^{-1}\frac{1}{2}r}$$

where $$\tanh z = \tfrac{1}{2}r.$$

Here again space contains an *infinite* number of cubic miles at any instant. Comparing this case with that of a flat space we see that, for the same function R and for the same co-ordinate-value r, there is here a greater volume than before. Such spaces are called *hyperbolic* or of *negative curvature* and are characterized physically by being *more 'voluminous'* than flat spaces and therefore, *a fortiori*, than spherical ones.

LUMINOSITY-DISTANCE, D

It has been shown [Equs. (1.3) and (1.4)] that the distance of a spiral nebula is estimated from its apparent magnitude by a formula which implies that the brightness of the nebula falls off with distance according to the inverse square law. We shall now prove [2] that distance in this sense is not the same thing as the distance, l, determined by measurements with rigid scales, in terms of which the red-shift-distance relation (4.15) was expressed. To show this we calculate the loss of energy in a train of light-waves between the moment they leave a point P_1 $(r_1, 0, 0)$ and the moment of their arrival at the origin P_0 $(0, 0, 0)$.

Considering first the loss of energy in a single photon corresponding to a wave of length λ, the energy of the photon at P_1 is hc/λ whilst at P_0 it is $hc/(\lambda + d\lambda)$, where h is Planck's constant. Hence

$$\frac{\text{Energy of photon emitted at } P_1}{\text{Energy of photon received at } P_0} = \frac{\lambda + d\lambda}{\lambda} = \frac{R_0}{R_1},$$

by equation (4.12). Again suppose that n photons are

emitted at P_1 in time dt_1. These are received at P_0 in time dt_0 where, by (4.10), $dt_1/dt_0 = R_1/R_0$. It follows that the number of photons emitted at P_1 in unit time is $n_1 = \frac{n}{dt_1}$ whilst the number received at P_0 is $n_2 = \frac{n}{dt_0}$ where n_1, n_2 are related by

$$n_2 = n_1\frac{dt_1}{dt_0} = n_1\frac{R_1}{R_0}.$$

Now the energy, E, of the radiation from P_1 which falls on unit area perpendicular to the radius-vector P_0P_1 at P_0 is

E = (number of photons arriving per unit time) $\times$ (energy per photon) $\times$ (solid angle subtended at P_1 by this area).

If therefore $d\omega$ denote the solid angle, we have

$$E = n_1\frac{hc}{\lambda}\frac{R_1{}^2}{R_0{}^2}d\omega.$$

With regard to $d\omega$, we remember that the particles P_1, P_0 have fixed co-ordinates (r, θ, φ). We can therefore define $d\omega$ to be equal to the inverse square of the radius of the sphere of co-ordinate radius r_1 centred at the origin, and measured by rigid scales at the *moment of receipt* of the light at P_0. It follows that $d\omega$ is $(-c^2)$ times the reciprocal of the coefficient of $(d\theta^2 + \sin^2\theta d\varphi^2)$ in the expression (4.8), calculated at the moment t_0. Hence

$$E = \frac{n_1hc}{\lambda}.\frac{R_1{}^2}{R_0{}^4}.\frac{(1 + kr_1{}^2/4)^2}{r_1{}^2}.$$

If we write

$$D = \frac{R_0{}^2}{R_1}\frac{r_1}{1 + kr_1{}^2/4}, \quad . \quad . \quad . \quad (4.19)$$

we have

$$E = \frac{n_1hc/\lambda}{D^2}$$

Thus if we have a number of similar sources of light of wave-length λ with different co-ordinates r_1, their apparent magnitudes will vary according to the inverse square of

the quantity D. For this reason D is called the *luminosity distance*, or *the distance corrected for red-shift and curvature*

On the other hand, the distance of the same source a the *moment of emission of the light* as it would be measure by rigid scales is

$$l = R_1 \int_0^{r_1} \frac{dr}{1 + kr^2/4} \quad . \quad . \quad . \quad (4.20$$

It is clear that l and D are equal only so long as (*a*) $r^2/$ is negligible compared with unity, (*b*) $R(t_1)$ is approximatel equal to $R(t_0)$. In these circumstances the formulae (4.19 and (4.20) become

$$D = R_0 r_1 = l.$$

THE N, δ RELATION

We now come to a formula which, after the approximat velocity-distance relation (4.15), is the most easily com parable with observation. It connects the number c particles lying in a shell of radii r_1 and $r_1 + dr$ with th increment of the Doppler shift as we pass from one sid of the shell to the other.

Every particle of our system has fixed co-ordinate (r, θ, φ) in the space-time (4.8) and the fluid composed c these particles is isotropic and homogeneous throughou all space at a given value of t. Consider for a momen the space in which the particles have fixed co-ordinate whose metric is

$$d\bar{s}^2 = \frac{dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2}{(1 + kr^2/4)^2}.$$

In this space, a volume determined by rigid scales an corresponding to co-ordinate-increments dr, $d\theta$, $d\varphi$ at th position (r, θ, φ) would be

$$dv = dl_1 . dl_2 . dl_3,$$

where

$$dl_1 = \frac{dr}{1 + kr^2/4}, \quad dl_2 = \frac{r d\theta}{1 + kr^2/4}, \quad dl_3 = \frac{r \sin \theta d\varphi}{1 + kr^2/4}.$$

The number of particles per unit volume must be th

same throughout this space. Let this number be α. Then in volume dv the number of particles is $n = \alpha dv$. But in actual space-time (4.8), the volume corresponding to dv is

$$dV = R(t)dl_1.R(t)dl_2.R(t)dl_3 = R^3(t)dv.$$

Hence
$$n = \alpha\frac{dV}{R^3(t)}.$$

Thus the number of particles, dN, lying inside the shell of radii r_1 and $r_1 + dr$ is

$$dN = \frac{\alpha}{R^3}\left\{\frac{4\pi R^3 r_1{}^2 dr}{(1 + kr_1{}^2/4)^3}\right\}$$

so that finally
$$dN = \frac{4\pi\alpha r_1{}^2 dr}{(1 + kr_1{}^2/4)^3} \quad . \quad . \quad . \quad (4.21)$$

To connect N with δ, we note that the Doppler shift observed at time t_0 at the origin from light emitted by any particle within the shell at time t_1, is [by (4.12)]

$$\delta = \frac{R_0}{R_1} - 1, \quad . \quad . \quad . \quad . \quad (4.22)$$

since the time taken by light to travel through the shell is negligibly small. For the motion of a light-ray we have, as always,

$$\int_0^{r_1}\frac{dr}{(1 + kr^2/4)} = c\int_{t_1}^{t_0}\frac{dt}{R(t)} \quad . \quad . \quad (4.23)$$

It is possible to expand the right-hand sides of equations (4.22) and (4.23) as series in $(t_0 - t_1)$ and thus to express r_1 as a function of δ. Using this value of r_1 in (4.21) we obtain a formula connecting $dN/d\delta$ and δ. The necessary calculations are long and complicated. They have been carried out by W. H. McCrea [2] and K. K. Mitra and yield, to the order δ^4,

$$(1 + \delta)\frac{dN}{d\delta} = \frac{4\pi\alpha c^3}{R_0'^3}\left\{\delta^2 + \frac{2}{R_0'^2}(2R_0''R_0 - R_0'^2)\delta^3 + \right.$$
$$\left.\frac{1}{12R_0'^4}\begin{pmatrix} 11R_0'^4 - 46R_0'^2R_0R_0'' + 45R_0^2R_0''^2 \\ -10R_0'R_0'''R_0^2 - 4c^2kR_0'^2 \end{pmatrix}\delta^4 + \,.\right\} \quad (4.24)$$

This is the formula we require. We shall find that by combining the observational formulae (1.5) and (1.11), it is possible to obtain an empirical relation corresponding to this theoretical formula.

Before proceeding, we note that the first approximation to (4.24) is

$$\frac{dN}{d\delta} = \frac{4\pi\alpha c^3}{R_0'^3}\delta^2,$$

and that this holds for all functions $R(t)$ and for any value of k.

EMPIRICAL N, δ RELATION

The empirical formula corresponding to (4.24) is deduced as follows. We have

$$\log_{10}\delta = 0{\cdot}2m - 4{\cdot}967 \quad . \quad . \quad . \quad (4.25)$$
$$\log_{10}N = 0{\cdot}501m - 2{\cdot}758 \quad . \quad . \quad (4.26)$$

Since the first formula has been established for values of m down to 17 only, whilst the second holds down to $m = 21$, some extrapolation of (4.25) is necessary before the formulae can be used together. There are two possibilities to be taken into account:

(A) The velocity-distance relation may, after $m = 17$, so completely alter its character that (4.25) is no longer even as good an approximation to the true relation as (4.26) is to the accurate $\log_{10} N$, m relation. In a universe whose most striking large-scale characteristic is uniformity such a sudden discontinuity appears highly improbable. But, of course, there is no observational evidence against such a hypothesis. If it is accepted, there is no possibility of using (4.25) and (4.26) as we propose to do for determining the curvature of space.

(*B*) Alternatively, bearing in mind the admittedly approximate character of (4.26), we may extrapolate the velocity-distance relation (4.25) to $m = 21$ *as it stands*. This is equivalent to regarding it as a good approximation to the truth down to this magnitude.

Eliminating m from (4.25) and (4.26) we obtain

$$N = 4{\cdot}57 \times 10^{9}\,\delta^{\frac{5}{2}} \quad . \quad . \quad . \quad (4.27)$$

Hence

$$\frac{dN}{d\delta} = 1{\cdot}14 \times 10^{10}\delta^{\frac{3}{2}} \quad . \quad . \quad . \quad (4.28)$$

Comparing this with (4.24), it appears (since $\delta < 1$) that we have here a *more rapid* increase of N with δ than is predicted by the first approximation

$$\frac{dN}{d\delta} \sim \delta^{2}$$

of the theoretical formula.

We can, however, account for (4.28) by means of approximations of higher order in (4.24). To do this it is first necessary to represent the function $\delta^{\frac{3}{2}}$, in the range corresponding to $m = 18{\cdot}47$ to $m = 21{\cdot}03$, by a power series in δ whose first term involves δ^{2}. We calculate from (4.25) the associated values

$$\left.\begin{array}{llllll} m = & 18{\cdot}47 & 19 & 19{\cdot}4 & 20 & 21{\cdot}03 \\ \delta = & {\cdot}05 & {\cdot}07 & {\cdot}08 & {\cdot}11 & {\cdot}17 \end{array}\right\} \quad (4.29)$$

where the values of m are those to which the five nebular counts were made. Assuming that

$$\delta^{\frac{3}{2}} = a_1\delta^{2} + a_2\delta^{3} + a_3\delta^{4} + a_4\delta^{5} + a_5\delta^{6}$$

and substituting for δ from (4.29), five equations are obtained giving the five constants a_1 to a_5. We obtain

$$\delta^{\frac{3}{2}} = \{6{\cdot}2529\delta^{2} - 40{\cdot}059\delta^{3} + 75{\cdot}542\delta^{4} + 177{\cdot}12\delta^{5} - 71{\cdot}275\delta^{6}\} \quad (4.30)$$

Dividing out here by the coefficient of δ^{2} and substituting in (4.28) we arrive at an expansion for $dN/d\delta$ whose first two terms are approximately

$$\frac{dN}{d\delta} = 1{\cdot}14 \times 6{\cdot}25 \times 10^{10}\{\delta^{2} - 6{\cdot}41\delta^{3} + \;.\;.\;.\;\} \quad . \quad (4.31)$$

CURVATURE OF SPACE

Comparing the coefficients of δ^2 and δ^3 in (4.24)* and (4.31) we obtain, after division of (4.24) by $(1 + \delta)$,

$$\frac{4\pi\alpha c^3}{R_0'^3} = 1{\cdot}14 \times 6{\cdot}25 \times 10^{10} \quad . \quad . \qquad (4{\cdot}32)$$

and

$$\frac{2}{R_0'^2}(2R_0''R_0 - \tfrac{3}{2}R_0'^2) = -\,6{\cdot}41 \quad . \quad . \qquad (4{\cdot}33)$$

To these we add the result (4.16), viz.

$$\frac{R_0'}{R_0} = 2 \times 10^{-17} \quad . \quad . \quad . \quad . \qquad (4{\cdot}34)$$

In addition to these we have the formulae (4.4) and (4.5) for the pressure and density of matter in the universe. The pressure is supposed to be due to the random motions of the nebulae, by which is meant their individual motions after the general motion of recession has been deducted. Since these 'proper-motions' are comparatively small, it is usual to assume that $p = 0$ in (4.4). Hence we have

$$0 = \Lambda - \left(\frac{R_0'}{R_0}\right)^2 - \frac{2R_0''}{R_0} - \frac{c^2k}{R_0^2} \quad . \quad . \qquad (4{\cdot}35)$$

$$8\pi\gamma\rho = -\Lambda + 3\left(\frac{R_0'}{R_0}\right)^2 + \frac{3kc^2}{R_0^2} \quad . \quad . \qquad (4{\cdot}36)$$

where

$$8\pi\gamma = 1{\cdot}674 \times 10^{-6}. \quad . \quad . \qquad (4{\cdot}37)$$

Putting $p = 0$ in (4.6) gives $\rho \sim 1/R_0^3$, so that if M_0 is the mass of an average nebula

$$\rho = \frac{\alpha}{R_0^3}M_0 \quad . \quad . \quad . \quad . \qquad (4{\cdot}38)$$

Using (4.32) and (4.38) we obtain

$$\rho = \frac{1{\cdot}14 \times 6{\cdot}25 \times 10^{10}}{4\pi c^3}\left(\frac{R_0'}{R_0}\right)^3 M_0$$

* Comparison of the coefficients of higher powers of δ in (4·24) and (4·31) merely yields the values of the derivatives R_0''', R_0^{iv}, etc., of R_0. These are of no interest in the calculation of the curvature of space.

which, with $M_0 = 2 \times 10^{42}$ gr. and R_0'/R_0 given by (4.34), yields

$$\rho = 3 \cdot 4 \times 10^{-30} \text{ gr./cm.}^3 \quad . \quad . \quad (4.39)$$

in agreement with the estimate of 10^{-30} gr./cm.^{3} given by Shapley. The equations (4.33) and (4.34) give further

$$\frac{R_0''}{R_0} = -3 \cdot 4 \times 10^{-34} \text{ sec.}^{-2} . \quad . \quad (4.40)$$

whilst on adding (4.35), (4.36), we have

$$\frac{2kc^2}{R_0^{\,2}} = 8\pi\gamma\rho - 2\left(\frac{R_0'}{R_0}\right)^2 + \frac{2R_0''}{R_0} \quad . \quad . \quad (4.41)$$

In this equation we substitute from (4.37), (4.39), (4.34) and (4.40). It becomes

$$\frac{2kc^2}{R_0^{\,2}} = 5 \cdot 7 \times 10^{-36} - 15 \times 10^{-34} \quad . \quad (4.42)$$

Clearly the second term is much larger than the first so that approximately we have

$$k = -1 \quad . \quad . \quad . \quad . \quad . \quad (4.43)$$

$$R_0 = 11 \times 10^{26} \text{ cm.} = 3 \cdot 5 \times 10^8 \text{ parsec.} \quad (4.44)$$

Lastly, from the equation (4.35), making use of (4.34), (4.40) and (4.42), we obtain

$$\Lambda = -10 \cdot 3 \times 10^{-34} \text{ sec.}^{-2} \quad . \quad . \quad (4.45)$$

A similar calculation, using (1.12) instead of (4.25), reveals that the general character of these results is not altered materially by using the maximum value of the constant term in the velocity-distance relation.

Summarizing the conclusions arrived at in this paragraph we may say that the nebular counts indicate a *hyperbolic space of radius* $3 \cdot 5 \times 10^8$ *parsecs*. The *rate of expansion* of the universe is being *retarded* and the value of the *cosmical constant is also negative*. At any instant space is infinite and the average *density of matter in the universe is of the order of* 10^{-30} *gr./cm.*^{3}. These conclusions must, however, be accepted with caution. It has already been pointed out that the nebular count formula is subject to large errors: we have now increased the uncertainty by

making the extrapolation denoted by (B) on page 6
In the opinion of Sir A. S. Eddington,[7] the probab
errors are so great that the nebular count material may l
regarded as consistent with a universe of very large *positi*
radius of curvature so that all space so far surveyed
effectively flat.

EINSTEIN AND DE SITTER UNIVERSES

There are two special cases of (4.8) which are of theore
cal and historical interest. The first case, associated wi
the name of Einstein, arises if

$$k = 1,\ R(t) = constant = R_e.$$

The metric of the Einstein universe is therefore

$$ds^2 = dt^2 - \frac{R_e^2}{c^2}\frac{dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\varphi^2}{(1 + r^2/4)^2} \quad (4.4$$

Using (4.4), (4.5) the values of the pressure and density a

$$8\pi\gamma p_e = \Lambda - \frac{c^2}{R_e^2}, \quad 8\pi\gamma\rho_e = -\Lambda + \frac{3c^2}{R_e^2}.$$

In the Einstein universe the radius, R_e, is further restrict
by the assumptions that the pressure of matter is zero a
that $\Lambda > 0$ so that

$$p_e = 0, \quad \Lambda = \frac{c^2}{R_e^2}, \quad 8\pi\gamma\rho_e = \frac{2c^2}{R_e^2} = 2\Lambda \quad . \quad (4.4$$

Thus R_e^2 is inversely proportional to the cosmical consta
and the density of matter is constant in time as well as
space. There is clearly no expansion and no velocit
distance relation.

The Einstein universe has a definite mass whose value

$$M_e = \int_0^\infty \frac{4\pi\rho_e R_e^3 r^2 dr}{(1 + r^2/4)^3} = 2\pi^2 R_e^3 \rho_e = \frac{\pi c^2}{2\gamma} R_e$$

so that the total amount of matter in this universe, as we
as its volume, are finite.

The de Sitter universe is obtained by putting, in (4.8

$$k = 0,\ R(t) = e^{\sqrt{\frac{\Lambda}{3}}t} \quad . \quad . \quad . \quad (4.4$$

The metric is now

$$ds^2 = dt^2 - \frac{e^{2\sqrt{\frac{\Lambda}{3}}t}}{c^2}(dr^2 + r^2d\theta^2 + r^2\sin^2\theta d\varphi^2).$$

Substituting (4.48) into (4.4) and (4.5), we obtain

$$p_s = 0, \ \rho_s = 0.$$

Strictly speaking there is neither matter nor radiation in this universe. It is, however, usual to regard the de Sitter universe as a limiting case in which the effect of matter and radiation on the metric has become negligible owing to the greatness of the expansion. The Doppler shift is given by (4.12) and (4.48) and is of amount

$$\delta_s = e^{\sqrt{\frac{\Lambda}{3}}(t_0 - t_1)} - 1.$$

The total volume of this universe is infinite.

The obvious defects presented by these two universes when compared with observation are the absence of a Doppler effect in the Einstein case and the 'emptiness' in the de Sitter.

THE TIME-SCALE

As an illustration of a problem in which the Einstein universe is involved, we consider the length of time during which the expansion of the universe has been taking place. Assuming, in the absence of wholly conclusive evidence to the contrary, that $k = +1$ and $p = 0$ we obtain from (4.6), as before,

$$\rho = \frac{\alpha M_0}{R^3} \quad . \quad . \quad . \quad . \quad . \quad (4.49)$$

By (4.5) we also have

$$3\left(\frac{R'}{R}\right)^2 = \frac{8\pi\gamma\alpha M_0}{R^3} + \Lambda - \frac{3c^2}{R^2} \quad . \quad . \quad (4.50)$$

If now we make the further hypothesis that the *initial* state of the universe corresponded to an approximately

Einstein condition, we should have the following initial values for the radius and density, viz.:

$$R = R_e, \quad \rho = \rho_e.$$

Hence by (4.47), (4.49)

$$\frac{8\pi\gamma\alpha M_0}{R_e^3} = \frac{2c^2}{R_e^2} = 2\Lambda.$$

We can therefore write the equation (4.50) in the form

$$\left(\frac{dR}{dt}\right)^2 = \frac{c^2}{3}\left\{2\frac{R_e}{R} + \frac{R^2}{R_e^2} - 3\right\}$$

$$= \frac{c^2}{3RR_e^2}(R - R_e)^2(R + 2R_e).$$

This equation can be integrated by the substitution $R = R_e(1 + x)$ to give

$$t = \frac{\sqrt{3}R_e}{c}\left[\log\{x + 2 + \sqrt{(x^2 + 4x + 3)}\}\right.$$

$$\left. + \frac{1}{\sqrt{3}}\log\left\{\frac{x + \sqrt{(x^2 + 4x + 3)} - \sqrt{3}}{x + \sqrt{(x^2 + 4x + 3)} + \sqrt{3}}\right\}\right] + \textit{constant} \quad (4.51)$$

for the time of expansion in terms of the radius of space. This solution was, historically, the first one used to account for the recession of the nebulae and is due to Lemaître [3] who originally published it in 1927.

The second term in (4.51) shows that when $x = 0$, $t = -\infty$, whilst the first term yields $x = +\infty$, $t = +\infty$. Theoretically, therefore, the universe starts expanding from an Einstein radius and takes an infinite time to acquire any value of the radius appreciably greater than R_e. But it is generally agreed that such a logarithmic infinity is unlikely to have any physical significance. Indeed, Eddington [4] has calculated that from the time when $R = 1{\cdot}5\ R_e$ to the present day, when $R'/R = 2 \times 10^{-17}$ sec.$^{-1}$, it is possible to allow 10^{10} years at most. Now the evidence provided by the physical conditions and the distribution of the stars in the Galaxy suggests an age of 10^{12} years

for these bodies. The expansion of the universe must therefore have been taking place for only one-hundredth part of the time during which the stars have been developing.

To overcome this paradox it has been suggested by de Sitter[5] that in the actual universe the function $R(t)$ is of such a character that the universe *contracts* during an infinite time, reaches a minimum radius and then starts expanding. We happen to be observing it in this latter stage. However, it turns out that the time from the moment of minimum radius to the present day is now only some 10^9 years. Stellar evolution, on this hypothesis, must therefore have been practically complete before the present expanding stage of the universe began. The shortness of the time-scale thus remains an outstanding problem of cosmological theory.

THE UNIVERSE AND THE ATOM [6]

It would take us too far afield to discuss the modifications introduced by Eddington into the quantum theory which bring it into harmony with general relativity. One of the results of his discussion, however, is the possibility of treating the theory of the Einstein universe first by means of general relativity and secondly by the generalized quantum mechanics of a system of particles in their fundamental state of lowest energy. The Einstein universe is also regarded as the initial state of the actual universe as was done in our discussion of the time-scale. Eddington then finds a formula for the number of particles in the universe in terms of the constant of gravitation, the masses of the proton and the electron, and the electronic charge. By means of this number, the mass of the initial state of the universe is obtained and its radius calculated. The latter is approximately

$$R_e = 4 \times 10^8 \text{ parsec.} \quad . \quad . \quad . \quad . \qquad (4.52)$$

It is also possible to show that the limiting velocity of recession of distant objects is

$$432 \text{ km./sec. per } 10^6 \text{ parsec.}$$

According, therefore, to Eddington the metric of the universe is that originally given by Lemaître [Equ. (4.51)] but with the addition of the particular initial value (4.52) for the radius. There is thus no need to appeal to nebular counts to determine the nature and amount of spatial curvature. And the fact that these counts suggest a hyperbolic universe, in which $k = -1$, is most probably due, in Eddington's opinion,[7] to the uncertainties of the observations.

A much more serious difficulty in the way of accepting Eddington's solution of the cosmological problem lies in the form which he gives to the quantum theory and on which the whole result turns. Considerable criticism of Eddington's ideas exists and comes most strongly from workers in the field of quantum theory itself.

CONCLUSIONS

Summarizing the results of this chapter, we may say that the expanding universes of general relativity, which assume that the matter in the universe is, on the average, evenly spread out in space, give a satisfactory account of the observed phenomena. They indicate that the volume of space so far surveyed is a small fraction of the whole. But whether this is due to space being hyperbolic and therefore infinite in extent or whether it is because space is spherical but of very large radius, are questions which cannot yet be answered with certainty. At most, it may be said that the balance of astronomical evidence is in favour of hyperbolic space. The length of time from the moment when the nebulae were closest together to the present day is of the order of 10^{10} years at most. This is a period one hundred times shorter than that indicated by the present state of stellar evolution.

Incidentally it has been found that the term ' distance ' in an expanding universe is ambiguous so long as the method of measurement is not specified. Distance in terms of measurements with rigid scales is not the same thing as luminosity-distance deduced from apparent magnitudes. This dependence of the meaning of distance on the

ocess of measurement we shall find emphasized in the
nematical theory of the universe to which we now proceed.

BIBLIOGRAPHY

H. P. Robertson, *Proc. Nat. Acad. Sci. Philad.*, **15**, 822, 1929.
L. P. Eisenhart, *Riemannian Geometry*, § 26, Princeton Univ. Press, 1926.
W. H. McCrea, *Zeits. f. Astrophysik*, **9**, 290, 1935.
G. Lemaître, *Ann. Soc. Sci. de Bruxelles*, **47** A, 49, 1927. Translated in *Mon. Not. Roy. Astr. Soc.*, **91**, 483, 1931.
A. S. Eddington, *Mon. Not. Roy. Astr. Soc.*, **90**, 668, 1930.
W. de Sitter, *Mon. Not. Roy. Astr. Soc.*, **93**, 628, 1933.
A. S. Eddington, *Relativity Theory of Protons and Electrons*, Chap. 14, Camb. Univ. Press, 1936.
A. S. Eddington, *Mon. Not. Roy. Astr. Soc.*, **97**, 156, 1937.
E. Hubble, *Astrophys. J.*, **84**, 517, 1936. This paper contains an analysis of the nebular-count material leading to a very small spherical universe.

CHAPTER V

THE KINEMATICAL THEORY OF THE UNIVERSE

IN the previous chapters the general theory of relativity provided us with a scheme of ideas which had already achieved success as a theory of gravitation. We found that this scheme, which accounted for the small-scale gravitational motions observed in nature, was equally capable of dealing with the structure of the whole universe. But a moment's reflection will convince the reader that the most striking phenomenon exhibited by the universe, the recession of the spiral nebulae, has very little resemblance to gravitational phenomena as exhibited in the motions of planetary systems, double stars, &c. It is therefore legitimate to inquire whether a theory of the universe can be constructed without an *a priori* appeal to a theory of gravitation. The problem which was set by E. A. Milne, and of which he gave one solution, was that of first building up a theory of the whole universe and then, if possible, of deducing from it the necessity of small-scale gravitational motion.

GENERAL PRINCIPLES

We shall, with E. A. Milne, abandon the greater part of the fundamental assumptions of general relativity and, in particular, the principle of covariance. This we do because it is well known in practice that certain coordinate systems are greatly superior to others. Nevertheless, the notions (i) that every event is specified by

ur co-ordinates, (t, x^1, x^2, x^3), the first of which has the mensions of time and the other three those of length,) that all events may be mapped in a Riemannian space-ne with metric

$$ds^2 = g_{44}dt^2 - \sum_{p,q=1}^{3} g_{pq}dx^p dx^q \quad . \quad . \quad (5.1)$$

ıd (iii) that the world-lines of light-rays are null-geodesics space-time, appear to be implicitly retained in kine-atical theory. The interval ds in (5.1) will now, how-er, have the physical meaning of the proper-time be-een two events only if these lie on the world-line of me specified observer who carries a clock. If this is ot the case, ds need not have a meaning in terms of iysical measurements.

We shall again assume that the nebulae of the actual iiverse are represented by particles which appear as ints in space. But we shall now suppose that these irticles are in motion relative to the co-ordinate system nployed, in such a way that, as t increases, every particle cedes from all the rest. It follows that, for some value t, a state of ' maximum closeness ' of the particles must ave occurred. For mathematical simplicity, we regard is state as being the actual coincidence of all particles at le position $x^1 = 0$, $x^2 = 0$, $x^3 = 0$ at $t = 0$ and we call le event with these co-ordinates the event O. In the ibsequent motion, we postulate that every particle moves irectly away from O. More exactly, if at the event (t, x^r), a particle has the four-dimensional velocity

$$\left(\frac{dt}{ds}, \frac{dx^1}{ds}, \frac{dx^2}{ds}, \frac{dx^3}{ds}\right)$$

ien the latter is parallel to the velocity with which the article left O. We are thus led at once to conclude that ie world-line of a particle is the path of an absolute arallelism in space-time which passes through O. All ie particles of the system, of course, move along the aths of the same absolute parallelism defined in terms of

some one set of fundamental vectors. Their equations of motion will therefore, by (2.32), be of the form

$$\frac{d^2x^k}{ds^2} + \Lambda^k_{pl}\frac{dx^p}{ds}\frac{dx^l}{ds} = 0 \qquad (k, p, l = 1, 2, 3, 4) \quad (5.2)$$

and, since we have abandoned the principle of covariance, there is no need to investigate *a priori* the question of the invariance of these equations under co-ordinate transformations. *In* (5.2), s *is, by definition, the proper-time of the particle tracing out the path.*

We have thus at the outset been led to abandon the principle of equivalence which states that the world-line of a material particle is *necessarily* a geodesic of space-time. We have substituted for it the notion that the world-line is orientated towards the peculiar event O. We shall nevertheless find that, by a special choice of metric and of absolute parallelism, it may happen that the world-lines of material particles are still the geodesics of space-time.

Having defined the motion of a single particle, we shall next (page 87) regard the whole system of particles as forming a fluid in motion. This fluid will be characterized by possessing, at each event P, density and momentum. We define these as follows. At the event P there is a particle of the set whose velocity is

$$\left(\frac{dt}{ds}, \frac{dx^r}{ds}\right) \qquad (r = 1, 2, 3).$$

The density of the fluid at P is then $\rho\frac{dt}{ds}$ and its momentum is $\left(\rho\frac{dx^1}{ds}, \rho\frac{dx^2}{ds}, \rho\frac{dx^3}{ds}\right)$. Lastly we postulate that this fluid obeys the equation of continuity defined by the metric.

The definition of the density-momentum vector and the continuity requirement correspond to the use of the Einstein gravitational equations in the expanding universe theory to define the state of matter in the universe.

THEORY OF EQUIVALENCE

Our first step must be to define the physical meanings of the co-ordinates (t, x^r) which we propose to use. We abandon the definitions of general relativity, as well as the notion of a rigid measuring-rod, and rely only on the idea of time.

The reason for doing this suggested by E. A. Milne is that human observers have an intuitive notion of the lapse of time, but not of spatial distance. The latter is a subsidiary notion derived from that of time. Without attempting to justify the philosophical implications of this doctrine, we may well admit that *measurements* of distance are reducible to measurements of time by means of the following device. Let there be an observer A, armed with a clock, and let him send out a light-signal to a distant point B, where it is immediately reflected and returned to A. Then A can regard the distance of B as a certain function of the times of dispatch and return of the signal as read on his clock.

Accepting this method of reducing space to time measurements, we can define a class of observers called by Milne *equivalent.** Two such observers A_0 and A_1 are defined as follows. Let s_0' be the proper-time of emission by A_0 of a light-signal which reaches A_1 at *his* proper-time s_1. Conversely, let s_1' be the proper-time of emission by A_1 of a signal which reaches A_0 at proper-time s_0. Suppose also that $+c$ is some parameter specifying the signal from A_0 to A_1 and $-c$ that from A_1 to A_0. Then A_0 and A_1 are *equivalent* observers if there exist relations of the form

$$F_0(s_0', s_1, c) = 0, \quad F_1(s_0, s_1', c) = 0 \quad . \quad . \quad (5{\cdot}3)$$

which are (a) independent of the particular light-signals used; (b) convertible into one another by interchanging s_0 and s_0', s_1 and s_1', and changing the sign of c.

* This use of the term ' equivalent ' has no relation to its use in general relativity where it figures in the ' principle of equivalence '.

A set of equivalent observers occurs when any observer of a group is equivalent in the above sense to all the other members of the group.

In what follows we shall limit ourselves to the following special problem. Firstly, we shall assume a metric of the form

$$ds^2 = \frac{dt^2 - \frac{1}{c^2}\{(dx^1)^2 + (dx^2)^2 + (dx^3)^2\}}{\left[1 - K^2\left\{t^2 - \frac{(x^1)^2 + (x^2)^2 + (x^3)^2}{c^2}\right\}\right]^2} \quad (5.4)$$

where the co-ordinates are undefined except for their dimensions. The constant c has the dimensions of velocity and the constant $K^2(>0)$ those of $(\text{time})^{-2}$. We also assume that the peculiar event O has the co-ordinates $(0, 0, 0, 0)$ so that near O the space-time is flat. Secondly, we shall choose the fundamental vectors of our parallelism so that at O the directions of these vectors shall be coincident with the directions of the co-ordinate axes. Thirdly, we shall assume that the observers are attached to the moving particles whose world-lines are paths through O. Thence we shall determine which observers are equivalent and what physical meanings can be given to the co-ordinates.

The foregoing choice of metric and of absolute parallelism is justifiable only *a posteriori* on the grounds that it leads to the most obvious generalization of Milne's original kinematical theory and that it also enables us to compare this theory most easily with general relativity. It is not difficult, however, to apply the method we shall use to the construction of kinematical theories in other space-times

WORLD-LINES OF OBSERVERS

We calculate first the finite equations of the observers' world-lines. These are given by the solutions of the differential equations (5.2) corresponding to paths passing through O. For simplicity we put

$$ict = x^4, \quad ics = \tau \quad . \quad . \quad . \quad . \quad . \quad (5.5)$$

$$q = -\log\left[1 - K^2\left\{t^2 - \frac{(x^1)^2 + (x^2)^2 + (x^3)^2}{c^2}\right\}\right]$$

$$= -\log\left[1 + \frac{K^2\{(x^4)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2\}}{c^2}\right] \quad (5.6)$$

so that the metric takes the form

$$d\tau^2 = e^{2q} \sum_{p=1}^{4} (dx^p)^2 \quad . \quad . \quad . \quad . \quad (5.7)$$

At O, the value of q is zero. Hence the fundamental vectors, which must be coincident with the co-ordinate-axes at that point, are

$$^{\alpha}h_p = \delta^{\alpha}_p e^q, \quad {}_{\alpha}h^p = \delta^p_{\alpha} e^{-q} \quad . \quad . \quad . \quad (5.8)$$

The paths are then

$$\frac{d^2x^k}{d\tau^2} + \frac{dx^k}{d\tau}\cdot\frac{dq}{d\tau} = 0 \qquad (k = 1, 2, 3, 4),$$

and their first and second integrals for paths passing through O are

$$\frac{dx^k}{d\tau} = \frac{A^k}{i}e^{-q},$$

and
$$x^k = \frac{c}{K}A^k \tanh\frac{K}{c}\frac{\tau}{i} \quad (k = 1, 2, 3, 4) \quad . \quad (5.9)$$

The constants of integration, A^k, are connected, in virtue of (5.7), by

$$\sum_{k=1}^{4} (A^k)^2 = -1 \quad . \quad . \quad . \quad . \quad . \quad (5.10)$$

Returning to the real variables t, s we put

$$\left(\frac{dt}{ds}\right)_0 = \frac{1}{\sqrt{(1 - V^2/c^2)}} = \beta,$$

$$\left(\frac{dx^r}{ds}\right)_0 = \frac{v^r}{\sqrt{(1 - V^2/c^2)}} = \beta v^r, \ (r = 1, 2, 3). \quad (5.11)$$

These constants therefore give the value of the four-dimensional velocity at O where, of course, $q = 0$. Hence we obtain

$$\frac{A^4}{i} = \beta, \quad A^r = \frac{v^r}{c}\beta \quad (r = 1, 2, 3),$$

and the relation (5.10) now becomes

$$\sum_{r=1}^{3} (v^r)^2 = V^2 \quad . \quad . \quad . \quad . \quad (5.12)$$

From (5.11) it follows that $\left(\dfrac{dx^r}{dt}\right)_0 = v^r$, $(r = 1, 2, 3)$.

We may therefore say that (v^1, v^2, v^3) are the components of the co-ordinate-velocity, V, with which the observer in question leaves the event O.

The finite equations of an observer's world-line are therefore

$$t = \frac{\beta}{K}\tanh Ks, \quad x^r = \frac{v^r\beta}{K}\tanh Ks, \quad (r = 1, 2, 3), \quad (5.13)$$

where we have chosen $s = 0$ at O.

We shall also require the following form for the first integrals of the world-lines. Consider

$$K^2\sigma^2 = K^2\left\{t^2 - \frac{(x^1)^2 + (x^2)^2 + (x^3)^2}{c^2}\right\}.$$

Substituting from (5.13), this gives $K\sigma = \tanh Ks$ along an observer's world-line. Hence, along this world-line,

$$\beta = t/\sigma, \quad v^r\beta = x^r/\sigma, \quad (r = 1, 2, 3),$$

and finally

$$\frac{dt}{ds} = \frac{t}{\sigma}(1 - K^2\sigma^2),$$

$$\frac{dx^r}{ds} = \frac{x^r}{\sigma}(1 - K^2\sigma^2), \quad (r = 1, 2, 3), \quad . \quad (5.14)$$

which gives the four-dimensional velocity of the observer at the event (t, x^r).

NULL-GEODESICS

Since equivalence involves the sending and receiving of light-signals, we require also the finite equations of the null-geodesics of the metric (5.4). In the notation of the last paragraph, we have $d\tau = 0$ along a null-geodesic. Hence if $\mu'(= i\mu)$ is an imaginary parameter varying along one of these curves, we have

$$\sum_{k=1}^{4}\left(\frac{dx^k}{d\mu'}\right)^2 = 0 \quad . \quad . \quad . \quad . \quad (5.15)$$

The non-zero Christoffel symbols are (there being no summation convention)

$$\{ml, m\} = -\{mm, l\} = \{ll, l\} = \frac{\partial q}{\partial x^l}$$

where $m \neq l = 1, 2, 3, 4$. Substituting these into the equations (2.22) of the null-geodesics and using (5.15), we have

$$\frac{d^2x^k}{d\mu'^2} + 2\frac{dx^k}{d\mu'}\frac{dq}{d\mu'} = 0, \quad (k = 1, 2, 3, 4),$$

the first integrals of which are

$$\frac{dx^k}{d\mu'} = \frac{a^k}{i}e^{-2q} \quad (k = 1, 2, 3, 4).$$

We return to real variables t, μ and we also put

$$a^4 = ic, \ a^r = p^r \quad (r = 1, 2, 3).$$

We then obtain

$$\frac{dt}{d\mu} = e^{-2q}, \quad \frac{dx^r}{d\mu} = p^r e^{-2q}, \quad (r = 1, 2, 3), \quad (5.16)$$

so that, by (5.15),

$$\sum_{r=1}^{3}(p^r)^2 = c^2 \quad . \quad . \quad . \quad . \quad (5.17)$$

From (5.16) it follows that $\frac{dx^r}{dt} = p^r$, $(r = 1, 2, 3)$, at all

points of space-time. Hence p^1, p^2, p^3 are the components of the constant co-ordinate-velocity c with which the null-geodesic is traced out. This co-ordinate-velocity is therefore identifiable with that of light.

The integrals of (5.17), giving a null-geodesic through the event $(t_0, x_0^1, x_0^2, x_0^3)$, are

$$t - t_0 = \frac{a\mu}{1 - b\mu}, x^r - x_0^r = p^r \frac{a\mu}{1 - b\mu} \quad (r = 1, 2, 3) \quad (5.18)$$

where a and b are constants given by

$$a = 1 - K^2\left\{t_0^2 - \frac{(x_0^1)^2 + (x_0^2)^2 + (x_0^3)^2}{c^2}\right\},$$

$$b = 2\frac{K^2}{c^2}\{- c^2 t_0 + x_0^1 p^1 + x_0^2 p^2 + x_0^3 p^3\}.$$

These are therefore the required finite equations of a null-geodesic.

EQUIVALENCE OF OBSERVERS

Consider now two observers A_0 and A_1 whose world-lines are respectively

$$t = \frac{\beta_0}{K} \tanh K s_0, \quad x^r = \frac{v_0^r \beta_0}{K} \tanh K s_0 \quad (r = 1, 2, 3) \quad (5.19)$$

and

$$t = \frac{\beta_1}{K} \tanh K s_1, \quad x^r = \frac{v_1^r \beta_1}{K} \tanh K s_1 \quad (r = 1, 2, 3) \quad (5.20)$$

where

$$\beta_0 = 1/\sqrt{(1 - V_0^2/c^2)}, \quad \beta_1 = 1/\sqrt{(1 - V_1^2/c^2)}.$$

Let A_0 emit a light-signal at the event (t_0, x_0^r) on his world-line, which meets A_1's world-line at the event (t_1, x_1^r). For definiteness suppose that $t_1 > t_0$ and $x_1^r > x_0^r$ $(r = 1, 2, 3)$. The two events must also lie on a null-

geodesic. Taking $\mu = 0$ at (t_0, x_0^r), we have by (5.18), (5.19) and (5.20)

$$\left.\begin{aligned} t_0 &= \frac{\beta_0}{K}\tanh Ks_0', \quad x_0^r = \frac{v_0^r\beta_0}{K}\tanh Ks_0', \ (r = 1, 2, 3) \\ t_1 &= t_0 + \frac{a\mu}{1 - b\mu} = \frac{\beta_1}{K}\tanh Ks_1, \\ x_1^r &= x_0^r + p^r\frac{a\mu}{1 - b\mu} = \frac{v_1^r\beta_1}{K}\tanh Ks_1, \ (r = 1, 2, 3) \end{aligned}\right\} \tag{5.21}$$

where s_0' is the time by A_0's clock at which he *emits* the signal and s_1 is the time by A_1's clock at which the latter *receives* it. From (5.21) we obtain, by eliminating μ, t_0, t_1, x_0^r, x_1^r, the *three* equations

$$\begin{aligned} &p^r(\beta_1 \tanh Ks_1 - \beta_0 \tanh Ks_0') \\ &\quad = (v_1^r\beta_1 \tanh Ks_1 - v_0^r\beta_0 \tanh Ks_0') \quad (r = 1, 2, 3). \end{aligned} \tag{5.22}$$

But, by definition, if A_0 and A_1 are equivalent there must be *one* relation only between s_0' and s_1, which is not dependent on the particular moment at which the signal is made. We can make A_0 and A_1 equivalent by putting

$$v_1^r = p^r\theta(V_1), \quad v_0^r = p^r\theta(V_0) \quad (r = 1, 2, 3).$$

To prove this, we square and add each set of three equations. We obtain

$$\theta^2(V_0) = \left\{\sum_{r=1}^{3}(v_0^r)^2\right\}\Big/\left\{\sum_{r=1}^{3}(p^r)^2\right\},$$

$$\theta^2(V_1) = \left\{\sum_{r=1}^{3}(v_1^r)^2\right\}\Big/\left\{\sum_{r=1}^{3}(p^r)^2\right\}.$$

Hence, using (5.12), (5.17),

$$\theta(V_0) = \pm V_0/c\,; \quad \theta(V_1) = \pm V_1/c \tag{5.23}$$

Choosing the upper signs, we have

$$p^r = \frac{v_0^r}{V_0}c = \frac{v_1^r}{V_1}c, \quad (r = 1, 2, 3) \tag{5.24}$$

so that the light-signal has a positive velocity in the coordinate system and is therefore travelling from A_0 to A_1,

as already assumed. The three equations (5.22) now reduce to the *single* equation

$$\sqrt{\left(\frac{1 - V_0/c}{1 + V_0/c}\right)} \tanh Ks_0' = \sqrt{\left(\frac{1 - V_1/c}{1 + V_1/c}\right)} \tanh Ks_1, \quad (5.25)$$

where s_0' is the proper-time of *emission* by A_0, and s_1 the proper-time of *receipt* of the light-signal by A_1. On the other hand, if we had taken the lower signs in (5.23) so that

$$p^r = -\frac{v_0^r}{V_0}c = -\frac{v_1^r}{V_1}c, \quad (r = 1, 2, 3) \quad . \quad (5.26)$$

the light-signal would have been travelling from A_1 to A_0. We should then have reduced the three equations (5.22) to

$$\sqrt{\left(\frac{1 + V_0/c}{1 - V_0/c}\right)} \tanh Ks_0 = \sqrt{\left(\frac{1 + V_1/c}{1 - V_1/c}\right)} \tanh Ks_1' \quad (5.27)$$

where now s_1' is the proper-time of *emission* by A_1, and s_0 that of *receipt* of A_0 of the light-signal. It is easily seen that (5.25) and (5.27), which correspond to (5.3), are symmetrical in the sense that either equation can be obtained from the other by interchanging s_0', s_0 and s_1', s_1 and simultaneously changing the sign of c.

Thus two observers are equivalent if the constants defining their world-lines and those defining the world-lines of light-signals joining them are given by (5.24) and (5.26). The proper-times of the two observers, which are measured by the clocks they carry with them, are then calibrated by the rules (5.25) and (5.27).

SETS OF EQUIVALENT OBSERVERS

Returning to the equations (5.19) and (5.20) of the world-lines of A_0 and A_1, we write these in the forms

$$\frac{x^1}{v_0^1/V_0} = \frac{x^2}{v_0^2/V_0} = \frac{x^3}{v_0^3/V_0} = V_0 t \quad . \quad . \quad (5.28)$$

and

$$\frac{x^1}{v_1^1/V_1} = \frac{x^2}{v_1^2/V_1} = \frac{x^3}{v_1^3/V_1} = V_1 t \quad . \quad . \quad (5.29)$$

On the other hand, the equation of the world-line of the light-signal which passes from A_0 to A_1 has the equation [by (5.18)]

$$c(t - t_0) = \frac{x^1 - x_0{}^1}{p^1/c} = \frac{x^2 - x_0{}^2}{p^2/c} = \frac{x^3 - x_0{}^3}{p^3/c}, \quad (5.30)$$

the event (t_0, x_0^r) being on the world-line of A_0.

The equations (5.28), (5.29) show clearly that the world-lines of A_0 and A_1 are straight lines through the origin $x^1 = 0$, $x^2 = 0$, $x^3 = 0$ of the spatial co-ordinates. And the equations (5.24), (5.26) reveal that these straight lines must be coincident. Lastly, (5.30) together with (5.24) or (5.26) shows that the light-signal also moves along the same spatial straight line. Thus, two equivalent observers move in space along the *same straight line* but with different (constant) velocities, the light-signals between the two observers also moving along this straight line.

The construction of a set of three or more equivalent observers is now easy. We have merely to secure that the observers of the set move with constant, but different, velocities along the same straight line in space, this line also passing through the origin of spatial co-ordinates.

PHYSICAL INTERPRETATION OF CO-ORDINATES

In the whole of the foregoing theory the only physical interpretation given to the co-ordinates has been a 'dimensional' one. The co-ordinate t has been specified no further than by requiring it to have the dimensions of time and the co-ordinates (x) have only been endowed with the dimensions of length. Indeed, a little reflection will convince the reader that the theory has so far been a matter of pure geometry and that the 'dimensions' of the co-ordinates have been introduced more for the purpose of providing us with a vocabulary* than because they were essential to the argument. We can now give a precise formulation of the co-ordinates of the event (t, x^r) in

* Thus in (5.14) we have been able to speak of 'velocity' instead of 'the rates of change of t, x^r with respect to s'.

terms of the proper-time measurements made by some observer, who possesses a clock but no rigid scales.

Let us regard the co-ordinate system (t, x^r) as that set up by an observer A_0 who gives himself the simplest spatial co-ordinates, $x^1 = 0$, $x^2 = 0$, $x^3 = 0$, and regards himself as at rest in the system. His world-line is therefore, by (5.19) with $V_0 = 0$,

$$t = \frac{1}{K} \tanh K s_0, \quad x^r = 0, \quad (r = 1, 2, 3) \,. \quad (5.31)$$

Let A_1 be a second observer equivalent to A_0. By the result of the previous paragraph, the world-line of A_1 must be a straight line through A_0, so that we may choose the x^1-axis of co-ordinates to lie along this line. Hence A_1's world-line is, by (5.13),

$$t = \frac{1}{K} \frac{\tanh K s_1}{\sqrt{(1 - V_1^2/c^2)}},$$

$$x^1 = \frac{V_1}{K} \frac{\tanh K s_1}{\sqrt{(1 - V_1^2/c^2)}}, \quad x^2 = 0, \quad x^3 = 0. \quad (5.32)$$

Suppose now that A_0 sends out a light-signal at his proper-time s_0' which travels to A_1, is immediately reflected there and returns to A_0 at proper-time s_0. If the proper-time of arrival at A_1 is s_1 $(\equiv s_1')$, we have, for the outward-going signal, by (5.25),

$$\tanh K s_0' = \sqrt{\left(\frac{1 - V_1/c}{1 + V_1/c}\right)} \tanh K s_1$$

and, for the returning signal, by (5.27),

$$\tanh K s_0 = \sqrt{\left(\frac{1 + V_1/c}{1 - V_1/c}\right)} \tanh K s_1.$$

Consider now the sum

$$\tanh K s_0 + \tanh K s_0'$$

and the difference

$$\tanh K s_0 - \tanh K s_0'.$$

We have

$$\tanh Ks_0 + \tanh Ks_0' = \frac{2 \tanh Ks_1}{\sqrt{(1 - V_1^2/c^2)}} = 2Kt \quad (5.33)$$

$$\tanh Ks_0 - \tanh Ks_0' = \frac{2V_1}{c} \frac{\tanh Ks_1}{\sqrt{(1 - V_1^2/c^2)}} = \frac{2Kx^1}{c} \quad (5.34)$$

where s_1 is A_1's proper-time of *receipt* as well as of *dispatch* of the signal. The time and space co-ordinates which A_0 assigns to A_1 are therefore

$$\frac{\tanh Ks_0' + \tanh Ks_0}{2K} \quad \text{and} \quad \frac{c(\tanh Ks_0 - \tanh Ks_0')}{2K},$$

respectively. They involve only the measured proper-times s_0', s_0 and the constants c and K. Since A_0 has no rigid scale he may *define* x^1, given by (5.34), to be the *distance* of A_1. We shall see later that the constant K can be found by observation so that the arbitrary element in (5.33) and (5.34) is the velocity c. If A_0 is an observer similar to terrestrial observers he may adopt for c the value 3×10^{10}. This fixes the scale of distance in his co-ordinate system in units of length he arbitrarily calls 'centimetres' and in units of time he calls 'seconds'.

Similarly the co-ordinates assigned by A_0 to any event (t, x^r) at which there is an equivalent observer, can be constructed out of the times of emission and receipt of light-signals. We must, however, also assume that A_0 has a theodolite with which to measure the angles between the directions in which the various equivalent observers lie. This is necessary in order that the observer may obtain the correct orientation of all events around him.

We shall need an alternative expression for the co-ordinates assigned to A_1, obtained by writing $s_1 = s_1'$ in (5.32) and then using (5.27) with $V_0 = 0$. The co-ordinates become

$$t = \frac{1}{K} \frac{\tanh Ks_0}{1 + V_1/c}, \quad x^1 = \frac{V_1}{K} \frac{\tanh Ks_0}{1 + V_1/c}, \quad x^2 = 0, \quad x^3 = 0 \quad (5.35).$$

In these formulae, s_0 denotes the *time of arrival* at A_0 of a light-ray emitted by A_1, so that t, x^1 refer to the moment of *emission* of the signal by A_1.

THE MOST DISTANT OBSERVER

If in (5.27) we put $V_0 = 0$ and $V_1 > c$, the proper-time of arrival at the observer A_0 of a light-signal emitted by A_1 is imaginary and devoid of physical significance. Hence c is the greatest permissible speed which an observer visible to A_0 may have. Putting $V_1 = c$ in (5.35) we obtain for the distance of the observer with this velocity,

$$x^1 = \tfrac{1}{2}\frac{c}{K}\tanh Ks_0.$$

This is therefore the maximum distance which the equivalent observer A_1 can have in order to be still visible to A_0 at proper-time s_0. It therefore constitutes the radius of the visible universe for the latter.

This result provides an additional interpretation for the velocity c which we have already identified with the velocity of the light-signals that equivalent observers dispatch to one another. We can now also regard it as the limiting velocity possessed by one observer who is still visible to another equivalent observer.

The foregoing method of assigning co-ordinates implies that the co-ordinates of an event are significant only if that event lies on the world-line of some observer. Any event, if such exists, which has never corresponded to a moment in the history of an observer has no physically significant co-ordinates.

RECIPROCITY OF METHOD OF ASSIGNING CO-ORDINATES BY TWO EQUIVALENT OBSERVERS

Consider again the observers A_0 and A_1, the former being at rest in the co-ordinate system (t, x) so that the world-lines of the observers are (5.31) and (5.32) respectively. If we transform to co-ordinates (t', x') by means of the formulae

$$t' = \frac{t - V_1 x^1/c^2}{\sqrt{(1 - V_1^2/c^2)}},$$

$$x'^1 = \frac{x^1 - V_1 t}{\sqrt{(1 - V_1^2/c^2)}}, \quad x'^2 = x^2, \quad x'^3 = x^3; \qquad (5.36)$$

and substitute in (5·32), A_1's world-line becomes

$$t' = \frac{1}{K}\tanh Ks_1, \quad x'^r = 0 \quad (r = 1, 2, 3) \,. \quad (5\cdot37)$$

Hence A_1 is at rest at the spatial origin of co-ordinates in the system (t', x'). On the other hand, A_0's world-line, (5·31), transforms into

$$t' = \frac{1}{K}\frac{\tanh Ks_0}{\sqrt{(1 - V_1^2/c^2)}},$$

$$x'^1 = -\frac{V_1}{K}\frac{\tanh Ks_0}{\sqrt{(1 - V_1^2/c^2)}}, \quad x'^2 = 0, \quad x'^3 = 0, \quad (5\cdot38)$$

so that A_0 is moving in this system with velocity $-V_1$. If in (5·38) we write $s_0 = s_0'$ and use formula (5·25) with $V_1 = 0$, $V_0 = -V_1$, we obtain for the co-ordinates assigned to A_0 by A_1, the values

$$t' = \frac{1}{K}\frac{\tanh Ks_1}{\sqrt{(1 + V_1/c)}},$$

$$x'^1 = -\frac{V_1}{K}\frac{\tanh Ks_1}{(1 + V_1/c)}, \quad x'^2 = 0, \quad x'^3 = 0. \quad (5\cdot39)$$

In these, s_1 is now the proper-time of *arrival* at A_1 of a signal *emitted* by A_0 at proper-time s_0'.

The formulae (5·37), (5·38) indicate that (t', x') have the same physical meanings for A_1 that (t, x) had for A_0. It also follows from (5·39) that the radius of the visible universe for A_1 at proper-time s_1 is $(c \tanh Ks_1)/2K$ which is precisely analogous to the value of the radius found by A_0.

Thus A_1 assigns co-ordinates and calculates the radius of the visible universe in terms of his proper-time s_1 by exactly the same types of formulae as A_0 does in terms of his proper-time s_0.

Finally, we have the important result that the co-ordinate systems in which A_0 and A_1 are respectively at rest are connected by the formulae of the Lorentz transformation, (5·36). We must, however, guard against the conclusion that this is a *necessary* property of the co-ordinate

systems used by equivalent observers. It only holds for equivalence of observers in space-times such as (5.4) which are conformal to flat space-time. A different initial choice of metric would not prevent us from defining equivalent observers, but the co-ordinate systems of such observers would not then necessarily be connected by Lorentz transformations.[1]

DOPPLER EFFECT

We have still to show that the constant K involved in the definitions of the co-ordinates has a physical significance. At the same time we can show that the velocity V_1 of an observer A_1 can be determined by measurements made by another equivalent observer A_0. Let A_0 again be at rest at the spatial origin of the co-ordinate system (t, x), let A_1 move along the x^1-axis and consider the Doppler shift measured by A_0 in the light he receives from A_1. By differentiating equation (5.27), with $V_0 = 0$, the Doppler shift found by A_0 at proper-time s_0, is

$$\delta = \frac{ds_0}{ds_1'} - 1 = \sqrt{\left(\frac{1 + V_1/c}{1 - V_1/c}\right)\frac{\operatorname{sech}^2 Ks_1'}{\operatorname{sech}^2 Ks_0}} - 1. \quad (5.40)$$

Suppose now that this Doppler shift is measured in the light-waves of a signal, originally sent out by A_0 at his proper-time s_0', which A_1 reflected at *his* proper-time s_1 $(= s_1')$ and which finally returned to A_0 at proper-time s_0. By (5.34) we have

$$\tanh Ks_1' = \tfrac{1}{2}\frac{c}{V_1}\left(1 - \frac{V_1^2}{c^2}\right)^{\frac{1}{2}}(\tanh Ks_0 - \tanh Ks_0').$$

Hence (5.40) becomes

$$1 + \delta = \left(\frac{1 + V_1/c}{1 - V_1/c}\right)^{\frac{1}{2}} \frac{\left\{1 - \frac{1}{4}\frac{c^2}{V_1^2}\left(1 - \frac{V_1^2}{c^2}\right)(\tanh Ks_0 - \tanh Ks_0')^2\right\}}{\operatorname{sech}^2 Ks_0} \quad (5.41)$$

In this equation there are two unknowns, V_1 and K, and three quantities directly observed, s_0', s_0 and δ. If there-

fore at some later time A_0 repeats the measurement of the Doppler shift and of the times of dispatch and return of a signal *sent to the same observer* A_1, he will obtain a second equation of the same form as (5.41) but with different values of s_0', s_0 and δ. These two equations then permit him to find V_1 and K in terms of directly observable quantities.

The constant K^2 in (5.4) clearly measures the departure of the metric from that of flat space-time and is therefore a measure of the curvature of space-time. The foregoing result demonstrates that, as in the expanding universes of general relativity, kinematical theory also predicts that the amount of the curvature of space-time can be found by observation.

We may summarize the results achieved so far by saying that the method of light-signals enables any observer to construct a co-ordinate system in which all equivalent observers can be located. The method also provides a definition of the distance of any one of these observers in terms of proper-time measurements alone. In this system the arbitrary element is the value of the velocity of light, c. This velocity is also that of the most distant observer visible to a given equivalent observer. Lastly, the method of light-signals, together with observations of the Doppler effect, provides a means of calculating the velocity of any specified observer and, at the same time, gives the value of the curvature of space-time.

SYSTEM OF PARTICLES

We now turn to the consideration of the whole system of equivalent observers each one of whom we regard as attached to a material particle. This system of material particles, as we have already said, will be supposed to constitute a hydrodynamic fluid of a simple type. Its state is characterized by a single four-dimensional vector, the components of which at any event (t, x^r) are the density of the fluid, corresponding to the time-component of the vector, and the three components of momentum, corresponding to the remaining components of the vector. In

classical mechanics such a fluid would be described [Equ. (3.11)] by an energy-tensor whose only non-zero components would be T^{44}, T^{4r} ($r = 1, 2, 3$). This particle-fluid we wish to identify with the system of the spiral nebulae.

The density-momentum vector describing the state of the fluid at every event (t, x^r) is assumed to be proportional to the four-dimensional velocity of the particle which has travelled from O to the event in question. If therefore the density-momentum vector be denoted by (j^4, j^1, j^2, j^3), j^4 being the density, then, by (5.14),

$$j^4 = \rho\frac{t}{\sigma}(1 - K^2\sigma^2),\ j^r = \rho\frac{x^r}{\sigma}(1 - K^2\sigma^2),\ (r = 1, 2, 3). \quad (5.42)$$

In these expressions ρ is a function of (t, x^r). It is to be determined by our second assumption with regard to the fluid, viz. that the vector (j^4, j^r) shall satisfy the equation of continuity (2.26) defined by the metric (5.4).

The determinant g for this case is

$$g = -\frac{1}{c^6}\left[1 - K^2\left\{t^2 - \frac{(x^1)^2 + (x^2)^2 + (x^3)^2}{c^2}\right\}\right]^{-8}$$

$$= -\frac{1}{c^6}(1 - K^2\sigma^2)^{-8} \quad . \quad . \quad . \quad (5.43)$$

In order not to have to deal with imaginary quantities, it is usual to multiply the equation (2.26) throughout by i so that it becomes

$$\frac{\partial(\sqrt{-g}V^p)}{\partial x^p} = 0.$$

With this slight alteration, the equation of continuity for our density-momentum vector in the space-time (5.4) takes the form

$$\frac{\partial}{\partial t}\left\{\rho\frac{t}{\sigma}(1 - K^2\sigma^2)^{-3}\right\}$$

$$+\sum_{r=1}^{3}\frac{\partial}{\partial x^r}\left\{\rho\frac{x^r}{\sigma}(1 - K^2\sigma^2)^{-3}\right\} = 0 \qquad (5.44)$$

Performing the partial differentiations, it appears that (5·44) is essentially a differential equation in which σ is the independent variable. This equation is

$$\frac{d}{d\sigma}\left\{\frac{\rho}{\sigma}(1 - K^2\sigma^2)^{-3}\right\} + \frac{4}{\sigma}\left\{\frac{\rho}{\sigma}(1 - K^2\sigma^2)^{-3}\right\} = 0.$$

The solution is

$$\rho = \frac{\alpha M_0}{c^3}\frac{(1 - K^2\sigma^2)^3}{\sigma^3}$$

where $\alpha M_0/c^3$ is the constant of integration in which M_0 is the mass of a particle and α is an arbitrary pure number. We assume for simplicity that all the particles have equal mass which we suppose expressed in grammes so that ρ is measured in gr./cm.^{3}. The density and momentum of the fluid at the event (t, x^r) are therefore

$$j^4 = \frac{\alpha M_0}{c^3} t \frac{(1 - K^2\sigma^2)^4}{\sigma^4},$$

$$j^r = \frac{\alpha M_0}{c^3} x^r \frac{(1 - K^2\sigma^2)^4}{\sigma^4}, \quad (r = 1, 2, 3). \quad (5·45)$$

Physically, this result means that an observer A_0 at rest at the spatial origin of the co-ordinate system (t, x^r), assigns the above values to the density and momentum of the particle-fluid in the neighbourhood of an equivalent observer at the event (t, x^r).

If now we consider any other equivalent observer, say A_1, who is moving with velocity V_1 along the x^1-axis, and transform the density-momentum vector to the system (t', x') in which A_1 is at rest, we obtain, using (5·36) and the law of vector transformation (2.6)

$$j'^4 = \frac{\alpha M_0}{c^3} t' \frac{(1 - K^2\sigma'^2)^4}{\sigma'^4}, \quad j'^r = \frac{\alpha M_0}{c^3} x'^r \frac{(1 - K^2\sigma'^2)^4}{\sigma'^4}, \quad (r = 1, 2, 3)$$

where $\quad \sigma'^2 = t'^2 - \frac{1}{c^2}\{(x'^1)^2 + (x'^2)^2 + (x'^3)^2\}.$

We conclude that the density-momentum vector has precisely the same form in terms of the co-ordinates employed by any observer of an equivalent set, provided

hat the co-ordinates are always assigned by the same rule of light-signals.

VELOCITY-DISTANCE RELATION

We have now to deduce the velocity-distance relation for the particle-fluid and compare it with the observed velocity-distance relation for the nebulae. We consider again the observer A_0, fixed at the spatial origin of co-ordinates. The three components of velocity of the fluid at the event (t, x^r) are, according to A_0,

$$\left(\frac{j^1}{j^4}, \frac{j^2}{j^4}, \frac{j^3}{j^4}\right) = \left(\frac{x^1}{t}, \frac{x^2}{t}, \frac{x^3}{t}\right).$$

Hence the fluid necessarily has a *radial velocity of recession* of amount r/t where $r^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$. *Thus for a given value of t, the velocity of the fluid varies directly with its distance from* A_0.

Since, however, the light-rays from events with the same t, but with different values of r, do not arrive simultaneously at A_0, this result does not correspond to the velocity-distance relation which A_0 *observes*. Again the time-co-ordinate t is not A_0's proper-time, *i.e.* it is not the time kept by A_0's clock. We have, therefore, to transform the velocity-distance relation to allow for these two facts.

Consider for simplicity the event $(t, x^1, 0, 0)$. The velocity of the fluid there is, by definition, equal to that of the equivalent observer at that event. Let this be V_1. To simplify further, let us suppose that the constant K is so small that we may write in equations (5.40) and (5.35)

$$\operatorname{sech}^2 Ks_1' = \operatorname{sech}^2 Ks_0 = 1, \qquad \frac{1}{K}\tanh Ks_0 = s_0. \quad (5.46)$$

The Doppler effect formula (5.40), from which V_1 is now determined, reduces to the well-known relation of special relativity

$$1 + \delta = \sqrt{\left(\frac{1 + V_1/c}{1 - V_1/c}\right)} \quad . \quad . \quad . \quad (5.47)$$

whilst the distance of the event is, by (5.35),

$$x^1 = \frac{V_1}{1 + V_1/c} s_0 \quad . \quad . \quad . \quad . \quad (5.48)$$

From the latter equation we obtain

$$V_1 = \frac{x^1}{s_0 - x^1/c} \quad . \quad . \quad . \quad . \quad (5.49)$$

From this equation it follows that, even for regions round the origin for which the approximations (5.46) are valid, the velocity of the fluid observed by A_0 at his proper-time s_0 does not vary directly with the distance but according to the law (5.49). As in the expanding universes of general relativity, the present theory gives a *linear velocity-distance relation only as a first approximation to a more complicated law.* Indeed, to obtain a linear relation from (5.49) it is obviously necessary to make a further assumption, additional to (5.46), and to assume that x^1 is so small that x^1/c is negligible over the regions of the fluid surveyed.

COMPARISON WITH OBSERVATION. TIME OF EXPANSION

The formula (5.49) may be regarded as giving the velocity V of *any* observer or particle of the fluid distant r from the origin of spatial co-ordinates at time s_0 registered by a clock at the origin. For the x^1-axis of co-ordinates may be supposed drawn in any arbitrary direction through the origin and the relation (5.49) still remains true. Hence

$$V = \frac{r}{s_0 - r/c}$$

and

$$\frac{dV}{dr} = \frac{s_0}{(s_0 - r/c)^2}.$$

Assuming that the approximations (5.46) hold for the actual universe in so far as it has been surveyed, and neglecting also the quantity r/c, the last equation becomes

$$\frac{dV}{dr} = \frac{1}{s_0}.$$

In this relation, s_0 is the proper-time since the observer at the origin left O, *i.e.* the proper-time since the expansion of the particle-fluid began. Using the same mean of the observed values as we did in establishing (4.16) we have

$$\frac{dV}{dr} = \frac{cd\delta}{dr} = \frac{3 \times 10^{10} \times 2{\cdot}2 \times 10^{-3}}{3{\cdot}1 \times 10^{24}} \text{ sec.}^{-1}$$

Hence $$s_0 = 1{\cdot}6 \times 10^9 \text{ years.}$$

Kinematical theory in the space-time (5.4) under the conditions (5.46) therefore gives a time of expansion of the order of 2×10^9 years at most. Short as was the time-scale provided by the expanding universes of general relativity, we have here a shorter one still. But too much importance cannot be attached to this result as we are not in a position to say to what extent it is dependent on our assumption that all the nebulae initially coincided. As far as it goes, however, it is a reason for preferring the expanding universe theory.

SPACE-TIME OF SPECIAL RELATIVITY

The conclusions of the last two paragraphs lead us at once to the consideration of the limiting case $K = 0$. The metric is now that of special relativity

$$ds^2 = dt^2 - \frac{1}{c^2}\{(dx^1)^2 + (dx^2)^2 + (dx^3)^2\} \quad (5{\cdot}50)$$

and the density-momentum vector reduces to

$$j^4 = \frac{\alpha M_0}{c^3}\frac{t}{\sigma^4}, \quad j^r = \frac{\alpha M_0}{c^3}\frac{x^r}{\sigma^4} \quad (r = 1, 2, 3). \quad (5{\cdot}51)$$

The formulae (5.47) to (5.49) are now exact and hold for all distances, whilst the co-ordinates assigned to the event $(t, x^1, 0, 0)$ by the stationary observer at the origin are

$$t = \frac{s_0}{1 + V_1/c}, \quad x^1 = \frac{V_1 s_0}{1 + V_1/c}. \quad (5{\cdot}52)$$

This case was studied in detail by E. A. Milne, who, however, used a quite different method of development

from the one we have adopted. The simplicity of the formulae (5·51), (5·52) makes them particularly suitable for comparison with observation. If they should prove inadequate, it is always open to us to use the more complicated scheme given by starting from the metric (5·4).

A striking contrast with general relativity appears when we remember that, in the latter theory, the space-time of special relativity corresponds to a complete absence of matter and radiation. We have now obtained a theory in which it has proved possible to combine flat space-time with the presence of matter.

THE N, δ RELATION

Restricting ourselves to the case (5·50) we shall now compare the observed nebular count formula with theory. We first transform (5·50) to polar co-ordinates and obtain, by (4·7),

$$ds^2 = dt^2 - \frac{1}{c^2}(dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2). \quad (5{\cdot}53)$$

In this co-ordinate system the density-momentum vector takes the form

$$j^4 = \frac{\alpha M_0}{c^3}\frac{t}{\sigma^4}, \quad j^1 = \frac{\alpha M_0}{c^3}\frac{r}{\sigma^4}, \quad j^2 = 0, \quad j^3 = 0; \qquad \sigma^2 = t^2 - r^2/c^2 \quad (5{\cdot}54)$$

which follows on using the vector-transformation law (2.6). The distance r and time-co-ordinate t associated with the event (t, r, θ, φ) by the observer A_0 at his proper-time s_0 are now

$$t = \frac{s_0}{1 + V/c}, \quad r = \frac{s_0 V}{1 + V/c} \quad . \quad . \quad (5{\cdot}55)$$

where V is the velocity of the equivalent observer at the event. Incidentally these expressions clearly demonstrate the spherical symmetry of the fluid round the origin.

Proceeding now to the determination of the theoretical rate of increase in the number of nebulae with Doppler

shift, we have the following preliminary results. From the equations (5·55) we obtain

$$\frac{V}{c} = \frac{r/c}{s_0 - r/c}$$

whence

$$1 + \frac{V}{c} = \frac{s_0}{s_0 - r/c}, \quad 1 - \frac{V}{c} = \frac{s_0 - 2r/c}{s_0 - r/c}. \qquad (5·56)$$

These yield, using (5·55) again,

$$t = s_0 - \frac{r}{c}, \quad \sigma^2 = t^2 - \frac{r^2}{c^2} = s_0\left(s_0 - \frac{2r}{c}\right). \qquad (5·57)$$

Writing $V = V_1$ in (5·47) and using (5·56) we obtain

$$(1 + \delta)^2 = \frac{s_0}{s_0 - 2r/c} \quad . \quad . \quad . \quad (5·58)$$

from which we deduce

$$r = \frac{cs_0}{2} \frac{\delta(2 + \delta)}{(1 + \delta)^2} \quad . \quad . \quad . \quad . \quad (5·59)$$

We can now calculate the number of nebulae lying in the shell of radii r and $r + dr$ which are simultaneously visible at time s_0 at the origin. We transform the vector (j^4, j^r) from the co-ordinate t to the 'co-ordinate' s_0 by means of (2.6), (5·55), and obtain

$$dN = \left(\frac{j^4}{M_0} \frac{ds_0}{dt}\right) 4\pi r^2 dr = \left\{\frac{\alpha}{c^3} \frac{t}{\sigma^4}\left(1 + \frac{V}{c}\right)\right\} 4\pi r^2 dr.$$

Eliminating V, t by (5·56) and (5·57), we have

$$dN = \frac{4\pi\alpha}{c^3} \frac{r^2 dr}{s_0(s_0 - 2r/c)^2} \quad . \quad . \quad . \quad (5·60)$$

Again, on differentiating (5·58), keeping s_0 fixed, we obtain

$$(1 + \delta)d\delta = \frac{s_0}{(s_0 - 2r/c)^2} \frac{dr}{c} \quad . \quad . \quad (5·61)$$

so that, introducing (5·61) into (5·60) and eliminating r by (5·59), we obtain finally

$$\frac{dN}{d\delta} = \pi\alpha\delta^2 \frac{(2 + \delta)^2}{(1 + \delta)^3} \quad . \quad . \quad . \quad (5·62)$$

As in the universes of general relativity, the first approximation is again

$$\frac{dN}{d\delta} \sim \delta^2.$$

The first three terms in the expansion of (5.62) are now, however,

$$\frac{dN}{d\delta} = 4\pi\alpha(\delta^2 - 2\delta^3 + \tfrac{13}{4}\delta^4 \ldots) \,. \quad (5.63)$$

Comparing this formula with (4.31), it follows that we have here a *too rapid* increase of nebulae with Doppler shift as compared with observation. But it is very satisfactory that the theory does predict an apparent outward increase in the number of nebulae of the type which the observed counts suggest. It will require much greater certainty with regard to the observed nebular-count formula before (5.63) can be definitely rejected as contrary to observation.

KINEMATICAL THEORY AND GENERAL RELATIVITY

Returning to the more general metric (5.4), let us transform the co-ordinates (x^1, x^2, x^3) to polars (r, θ, φ) by (4.7) and then transform t, r to T, z by

$$t = \frac{1}{K}\frac{1+z^2/4}{1-z^2/4}\tanh KT, \quad r = \frac{c}{K}\frac{z}{1-z^2/4}\tanh KT. \quad (5.64)$$

We obtain

$$ds^2 =$$

$$dT^2 - \frac{1}{4K^2}\sinh^2 2KT\frac{(dz^2 + z^2 d\theta^2 + z^2\sin^2\theta d\varphi^2)}{(1-z^2/4)^2}. \quad (5.65)$$

Comparing this with (4.8) it follows that the metric (5.4) in terms of the co-ordinates (T, z, θ, φ) is that of an expanding universe of general relativity in which

$$R(T) = \frac{c}{2K}\sinh 2KT, \qquad k = -1$$

so that space is hyperbolic. Again, we can without loss of generality regard the observers A_0 and A_1, whose

world-lines are given by (5.31) and (5.32) as two *typical* equivalent observers. In terms of (T, z, θ, φ) A_0's world-line becomes

$$\tanh KT = \tanh Ks_0, \quad z = 0, \quad \theta = 0, \quad \varphi = 0,$$

whilst A_1's is, writing $r = x^1$ in (5.32),

$$\tanh KT = \tanh Ks_1, \quad z = 2\frac{\{1-\sqrt{(1-V_1^2/c^2)}\}}{V_1/c}, \quad \theta = 0, \quad \varphi = 0.$$

Hence A_0 and A_1 have *fixed* co-ordinates (z, θ, φ) and the co-ordinate T measures the *proper-time of both.* But referring to (4.2) this means that the world-lines of A_0 and A_1 are *geodesics* of the hyperbolic expanding universe of general relativity with metric (5.65). Our two equivalent observers are therefore a pair of 'relativistic' observers who have lost their rigid scales and have fallen back on the method of light-signals for the determination of distances, &c. Many investigators [2] have pointed out that, when the co-ordinate systems of equivalent observers are related by Lorentz transformations, kinematical theory bears a strong resemblance to the theory of hyperbolic universes. From the present point of view, this arises from the fact that in such universes geodesics and paths of certain absolute parallelisms happen to be identical.

COSMOLOGICAL PRINCIPLE

We have found that the expressions for the co-ordinates in terms of proper-times, the radius of the visible universe, and the density-momentum vector preserve *the same formal expressions* under Lorentz transformations. There is therefore invariance of form, but not, of course, of value, on transforming from the co-ordinates of one observer to those of an equivalent observer. This is expressed by E. A. Milne in a *cosmological principle* which states that:

'*The development of the universe appears to be the same for each observer of an equivalent set, every one of whom assigns co-ordinates by the same method.*'

In Milne's presentation of the theory this principle is made the starting-point and an attempt is made to build up the whole theory from it. But it does not prove pos-

sible to advance very far without additional hypotheses which appear to be equivalent to the assumptions that all events are plotted in the flat space-time of special relativity, that distances are then assigned by the method of light-signals and that every event is orientated towards the special event we have called *O*. This is a much more specialized state of affairs than that contemplated by the cosmological principle. The latter corresponds roughly to the principle of covariance in general relativity, since it indicates the *type* of formulae we are to use in the description of the universe, viz. such formulae as are invariant on transforming from the co-ordinates of one equivalent observer to those of another. Like the principle of covariance, the cosmological principle is too general to carry us very far without supplementary assumptions.

GRAVITATION

It still remains to discuss the meaning of gravitation in kinematical theory. There is general agreement amongst investigators that gravitation is connected with the existence of accelerated particles of matter in the universe and that the accelerations of such particles must, in the first approximation, reduce to the Newtonian inverse square law of attraction. But there is considerable disagreement as to what *kind of acceleration*, reducible in this way, is to be accounted gravitational.

The most determined attack on this problem has been made by E. A. Milne.[3] Restricting himself to flat space-time and interpreting the cosmological principle as meaning invariance under Lorentz transformations, this author calls the system of equivalent particles we have been studying the 'fundamental' system. Superimposed on this, he considers a general 'statistical' system of particles having the following properties: (1) Each particle of the statistical system has an acceleration, the expression of which is invariant under the Lorentz transformations connecting the co-ordinate systems of the observers travelling with the fundamental particles. (2) The development of the statistical system as a whole appears to be the same when viewed

from any fundamental particle. (3) The particles of the statistical system are not themselves equivalent in the sense of kinematical theory. These hypotheses yield for the acceleration of a particle of the statistical system the law

$$\frac{d^2x^r}{dt^2} = \left(x^r - t\frac{dx^r}{dt}\right)\frac{Y}{\sigma^2}G(\xi) \quad (r = 1, 2, 3) \quad (5.66)$$

where

$$Y = 1 - \frac{1}{c^2}\sum_{r=1}^{3}\left(\frac{dx^r}{dt}\right)^2, \; \xi = \frac{Z^2}{Y\sigma^2}, \; \sigma^2 = t^2 - \frac{1}{c^2}\sum_{r=1}^{3}(x^r)^2,$$

and

$$Z = t - \frac{1}{c^2}\left(\sum_{r=1}^{3} x^r\frac{dx^r}{dt}\right).$$

The function $G(\xi)$ is arbitrary and defines the particular statistical system envisaged. It must, however, be said that the rules (1), (2), (3) require only that G should be a function of the Lorentz invariants σ^2, Z^2/Y. The particular combination ξ follows from the *additional* assumption that the law (5.66) must not involve any arbitrary constant of the dimensions of length or of time. The combination ξ of σ^2 and Z^2/Y is the only one giving such a dimensionless argument for G.

If then the acceleration (5.66) is really the gravitational law of motion in the universe, it should be possible to deduce from it the value of the gravitational constant γ, known to be involved in this law. Consider for definiteness the fundamental observer A_0 at rest at the spatial origin of the co-ordinates (t, x) whose world-line is, by (5.31) with $K = 0$,

$$t = s_0, \; x^r = 0 \; (r = 1, 2, 3).$$

The co-ordinate t is therefore identical with A_0's proper-time. Milne then shows that the motion (5.66) of a statistical particle near A_0 appears to this observer to be an inverse square law acceleration *provided that γ varies directly with A_0's proper-time t (or s_0).* Since A_0 is any observer chosen arbitrarily, it follows that this result will

be true for any equivalent, or fundamental, observer in terms of co-ordinates and proper-time of the latter. The constancy of γ, assumed in Newtonian and Einsteinian gravitational theory, must therefore be an illusion. According to every fundamental observer, this quantity is proportional to the time since the expansion of the universe began.

Since this conclusion is not entirely acceptable, Milne has tried to avoid it by showing that the law (5.66) can be reduced to an inverse square acceleration with a *constant* value of γ by a change of proper-time variable. Effectively Milne argues that A_0, at any instant t_0 of his proper-time, uses a 'time' $\tau = t_0 \log (t/t_0) + t_0$ for the measurement of gravitational accelerations. It can be shown that in terms of τ the law of motion (5.66) reduces to a strictly Newtonian inverse square law acceleration in the first approximation.

Some objections to this scheme are the following. In the first place, the presence of the particles of the statistical system is hypothetical: the existence of the fundamental system in no way implies the existence of the accelerated particles also. A universe full of matter might therefore exist in which there would be no naturally occurring gravitational phenomena. In the second place, the argument by which G is made to depend on ξ alone is arbitrary. The cosmological principle indicates that G is a function of $\sigma^2/\sigma_0{}^2$ and $Z^2/\tau_0{}^2 Y$ where $\sigma_0{}^2$ and $\tau_0{}^2$ are 'cosmical constants' whose physical significance is not very clear. Lastly, the hypothesis that an observer has one time-scale for cosmical, and another for gravitational, phenomena seems to disagree with the practice of terrestrial observers.

Other investigators [4] have, on the one hand, rejected Milne's narrow interpretation of the cosmological principle in terms of Lorentz transformations and, on the other hand, have frankly accepted the existence of γ as a constant of nature whose value cannot be deduced mathematically. The kinematical theories so constructed in curved space-times have proved compatible with the laws of gravitation of Newton and of Einstein.

In conclusion we can only say that the problem of gravitation in kinematical theory still awaits a completely satisfactory solution.

CONCLUSIONS

Summarizing the results of this chapter we may say that kinematical theory gives a simpler interpretation of co-ordinates than does general relativity. It accepts the expansion of the universe as a fact and therefore does not suggest contraction as an alternative possibility and it accounts for the recession of the nebulae with a velocity proportional to distance but in the same approximate manner as is found in the expanding universe theory. The time-scale on this theory is the short one of some 10^9 years. Kinematical theory definitely predicts an increase in the number of nebulae with Doppler shift of the same type as is observed. But the problem of gravitation remains largely unsolved although it is true that all possibilities in this connexion have not yet been investigated. Even if a gravitational theory comparable with that of general relativity is never attained, yet kinematical theory will have served the important purpose of showing that the recession phenomenon is essentially distinct from the gravitational phenomena in the universe.

BIBLIOGRAPHY

1. A. G. Walker, *Proc. Lon. Math. Soc.*, Ser. 2, **42**, 90, 1937.
H. P. Robertson, *Astrophys. J.*, **82**, 284, 1935.
2. W. O. Kermack and W. H. McCrea, *Mon. Not. Roy. Astr. Soc.*, **93**, 519, 1933.
H. P. Robertson, *Zeits. f. Astrophysik*, **7**, 153, 1933.
A. G. Walker, *Mon. Not. Roy. Astr. Soc.*, **95**, 263, 1935.
3. E. A. Milne, *Proc. Roy. Soc., London*, Ser. *A*, **154**, 22, 1936; **156**, 62, 1936; **158**, 324, 1937.
4. G. C. McVittie, *Zeits. f. Astrophysik*, **10**, 382, 1935.
H. P. Robertson, *Astrophys. J.*, **83**, 187 and 257, 1936.

INDEX

Printed in Great Britain by
Butler & Tanner Ltd.,
Frome and London